

Universities confront changing markets

Tertiary education is a competitive business with growing export potential. A recent report highlights the importance of postgraduate training, but governments are undermining universities' abilities to compete in an expanding market.

IMAGINE a young and ambitious businesswoman, with a degree in computing and with several years' employment in a growing company. She is convinced that both she and her employer would benefit if she studied the application of computers applied to commerce and industry. She has to balance family and work commitments, so cannot simply move to an appropriate college. But she plugs into the World Wide Web, clicks into a search engine, specifies "online education", and discovers that the Open University can provide her with an online Masters course. Her employer will pay the full costs of the course, she is prepared to put in the hours needed to gain a significant qualification without sacrificing other commitments, the university will gain another customer, and no government has to subsidize the deal.

That scenario (which is possible today) represents a glimpse of a burgeoning market for higher education institutions, and an increasingly attractive one: taught courses for postgraduates, as distinct from research. The students are, increasingly, part-time and mature, and have access to private finance, whether their own or their employer's. Postgraduate science is dominated by research-based studies, but advanced taught courses, delivered locally or at a distance, are becoming increasingly saleable commodities. In a report published in the United Kingdom last week, Professor Martin Harris and colleagues provide the first overview of postgraduate education across all disciplines in British universities (see page 266). The report highlights the differences in support mechanisms and market potential between postgraduate taught and research degrees, showing that the growth of the former is now the fastest expanding component of university activity.

But this growth has been stimulated mainly by the need for universities to find money. Government resources for undergraduate education have diminished while undergraduate numbers have, until recently, grown. And government funding for research is tightly constrained, with support for infrastructure in steep decline. In developing new teaching courses while under such strain, there is a real risk that universities will give less value than is acceptable. The Harris report therefore provides a useful service by focusing on constraints and opportunities in postgraduate teaching and research. It is right to emphasize the primary role of individual institutions rather than national agencies in establishing the diversity of taught courses that can be supplied, while insisting that customers should not be confronted (as they are now) with a confusing plethora of course titles, with inadequate description of just what will be delivered. But it is realistic also in emphasizing the need for institutions to recover progressively more of the full cost of taught courses from the students themselves.

Guaranteeing quality will be crucial to this growing export industry, in the United Kingdom, in Australia (see page 265), and elsewhere. In the traditional university model, research has the added value of underpinning teaching quality. The Harris report accepts the nontraditional (in the United Kingdom) but increasingly realistic view that postgraduate teaching and research are distinct and separable activities. Controversially, it states that government support for postgraduate research should be selectively based on an institution's achievements in the periodic research assessment exercise carried out by higher education funding councils. It is right, however, to insist that funds for undergrad-

uate education should be protected, and that postgraduate taught courses should also be targeted for support by the funding councils but supplied within a more market-led and commercial approach.

But there is a question begged in all of this: is the system as a whole in good health? For an answer one has to refer instead to the much more critical document, *Research Capability of the University System*, produced last month by the UK National Academies Policy Advisory Group (see *Nature* 380, 571; 1996). Its identification of "tensions and irreconcilable contradictions" provides salutary warnings. Shortcomings in the research assessment exercise, overemphasis by research councils on technology foresight, obstacles to interdisciplinary and innovative research are just some of the problems highlighted. Unless they are addressed by the government, the potential benefits of the Harris report's recommendations on research training will be undermined. And unless undergraduate training receives the government support it requires, the market-led teaching of postgraduates will also be damaged. In both respects, the United Kingdom's ability to realize what it can offer postgraduate students is under threat. □

Treaty on the brink?

India should not threaten the success of the Comprehensive Test-Ban Treaty in its quest for disarmament.

INDIA deserves some credit for consistency and a lack of hypocrisy in nuclear affairs. It refused to sign the Non-Proliferation Treaty (NPT) and exploded a nuclear device in 1974. Provocatively, the three weapons states negotiating the treaty — the United States, Russia and the United Kingdom — were trying to limit nuclear proliferation while continuing to develop their own arsenals. India now suspects that while the Comprehensive Test-Ban Treaty (CTBT) under negotiation at Geneva (see page 267) will prevent weapons proliferation, it will not commit the nuclear powers to disarming. But in seeking such commitments in advance, India is asking too much of the real world.

A test-ban treaty was the price that the nuclear powers paid for last year's agreement among nonnuclear states on an indefinite extension of the NPT. Had the nuclear enthusiasts in the Pentagon got their way in demanding that the ban exempt small nuclear explosions, India would have been right to cry foul. But President Bill Clinton adhered instead to independent scientific advice which said such tests were not needed to guarantee the safety of the stockpile. All five nuclear powers now agree, with the result that the treaty under discussion at Geneva is more unequivocally comprehensive than anyone could have hoped for.

Agreement on a test ban is not an event in itself, but the beginning of a long process. By freezing the gap between the weapon haves and have-nots, a test ban will create better conditions for further talks on disarmament. But without a CTBT the NPT will be weaker, and the risk that dictators will obtain — and use — nuclear weapons will be greater. It would be a shame if India, which called for a test ban as long ago as 1955, should be the one to scuttle the ship as it pulls into port. □



Box of bones 'clinches' identity of Piltdown palaeontology hoaxer

London. A trunk discovered under the roof of London's Natural History Museum appears to have provided vital evidence allowing the Piltdown fraud — one of the most successful hoaxes in scientific history — finally to be put to rest.

The canvas travelling trunk is marked with the initials of Martin A. C. Hinton, a curator of zoology at the museum at the time of the fraud. It contains bones stained and carved in the same way as the Piltdown fossils and associated artefacts.

The discovery is the first solid evidence in the case after decades of speculation. As Brian Gardiner, professor of palaeontology at King's College, London, is due to announce in his presidential address to the Linnean Society in London tomorrow (24 May), it appears to identify Hinton unequivocally as the hoaxer.

The story of what was to turn into the longest-running parlour game in the history of palaeoanthropology began in 1912 when Charles Dawson, a lawyer and antiquary, unearthed human skull and jaw fragments and primitive artefacts at a gravel pit at Piltdown, Sussex, in the south of England.

The find caused a sensation. The Piltdown skull seemed remarkably advanced, given the great age indicated by the bones of archaic forms of fossil mammal characteristic of Pliocene deposits. But the jaw seemed very primitive, and almost ape-like. Both tied in with the prevailing views of human ancestry, namely that humanity was the culmination of a very ancient lineage, and that the first modern human feature to emerge was the enlarged brain.

Several people immediately suspected fraud. But many considered the skull to be an immensely important find, in particular Arthur Smith Woodward, keeper of palaeontology at the museum.

Subsequent research showed humanity evolved quite differently. The human lineage turns out to be relatively young, and expansion of the brain took place relatively late in human history. Piltdown Man was increasingly seen as an aberrant offshoot.

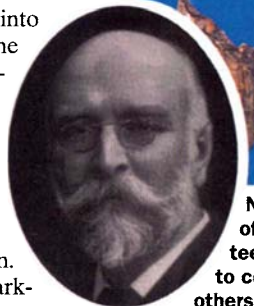
Nevertheless, Smith Woodward remained convinced of the authenticity of the skull until the day he died; his book *The Earliest Englishman*, published in 1948, was dictated on his deathbed. But just five years later, chemical analyses by Kenneth P. Oakley of the museum showed that all the artefacts were of recent date. The skull came from a

modern human, the jaw from an orang-utan.

The accompanying mammal fossils, it emerged, had been 'planted' to give the human fossils an authentic context. All the objects had been carefully stained and abraded to appear old. Artefacts such as a piece of elephant bone carved to look like a cricket bat (a fitting accoutrement



New evidence, old bones: the contents of a forgotten trunk included bones and teeth, stained using a technique that was to convince Smith Woodward (left) and others that the Piltdown skull was authentic.



for the 'first Englishman') were seen as part of an elaborate joke. But who was the joker, and why?

Dawson was the prime suspect. But as he died in 1916, he could hardly be confronted with the evidence. Indeed, since 1953, virtually everyone connected with Piltdown Man has come under suspicion as the hoaxer, from the distinguished anatomists Arthur Keith and Grafton Elliot Smith to Teilhard de Chardin, the palaeontologist priest, and even Sir Arthur Conan Doyle, the creator of Sherlock Holmes, who lived nearby and is known to have visited the site.

Hinton has not escaped previous suspicion. Writing in *The Common but Less Frequent Loon and Other Essays*, published this year, Keith S. Thomson suggested that Dawson initially perpetrated the fraud, but that it was immediately spotted by Hinton. But Thomson claimed that because Hinton was afraid of exposing his boss, Smith Woodward, to ridicule, he sought to expose the hoax by planting increasingly ridiculous fossils, such as the 'cricket bat', at the site in order to scare Dawson and expose him as a fraud in the eyes of Smith Woodward.

But the new evidence contained in Hin-

ton's trunk disproves this scenario. It now turns out that all the Piltdown remains were stained with the same chemical recipe, one that was invented by Hinton. The evidence appears to identify Hinton as the sole fraudster — and Dawson as his unwitting dupe.

The trunk came to light in the mid-1970s, when contractors were clearing the loft space in the southwest tower of the museum before maintenance work was carried out on the roof. It came to the notice of Andrew Currant, a researcher at the museum specializing, as did Hinton, in fossil rodents.

The trunk contained hundreds of vials of rodent dissections — described by Currant as "quite macabre". But at the bottom lay a hidden treasure: a collection of carved and stained pieces of fossil hippopotamus and elephant teeth, as well as assorted bones, that looked as if they belonged in the Piltdown collection.

Realizing the potential importance of what could be the 'smoking gun' to one of the great hoaxes of the century, Currant mentioned the existence of the trunk to Gardiner, who had been on the Piltdown trail since the hoax was first exposed in 1953, and was already certain from circumstantial evidence that it led to Hinton's door.

The trunk was the final piece of the puzzle — and the clinching solid evidence — that Gardiner needed to establish his case. He and Currant have spent the last few years studying its contents and reanalysing the Piltdown collection. As Gardiner is due to announce to the Linnean Society tomorrow, they are now certain that Hinton was the sole author of the fraud.

Hinton himself, who published widely on many aspects of zoology and palaeontology, was something of a prodigy. In 1899, at the age of sixteen, he published a paper showing how fossils in river gravels would be impregnated with oxides of iron and manganese, staining them a characteristic chocolate-brown colour.

Oakley's analyses showed that the Piltdown fossils were enriched in iron, as one would expect if they were genuinely old — even though their recent age was indicated by other factors. But Oakley did not look for manganese. Crucially, analyses of the contents of Hinton's trunk by Currant and Gardiner show that they are enriched in iron as well as manganese — in the same proportions as in the Piltdown specimens.

The Piltdown fossils — with one ►

► important exception — as well as the contents of Hinton's trunk, are also enriched with chromium. This appears to have been a result of the staining process.

Before staining, Hinton would have used chromic acid as an ingredient in a recipe to turn the apatite (the mineral component of bone) to gypsum. This process would have etched the bone surface, making it easier for the manganese and iron oxides to penetrate the specimens. But traces of chromium would remain.

The one exception was the orang-utan jaw. This could not be etched because it contained two teeth, and acid-etching of the teeth would have been a clear sign of a forger at work. Hinton was therefore careful not to treat these in the same way. The teeth were lightly stained so as not to risk etching, and an isolated canine tooth was painted further with paint (possibly burnt umber) rich in manganese and iron.

Gardiner and Currant's suspicions about Hinton's difficulties with teeth received support in 1991 after Gardiner contacted Robert J. G. Savage, then professor of geology at the University of Bristol, telling him of Currant's discovery of the Hinton trunk. Savage had been the executor of Hinton's estate — a considerable task, given that Hinton was a lifelong hoarder.

Savage sent Gardiner some glass tubes from Hinton's hoard. These contained eight human teeth that had been stained in various ways. The teeth, together with the con-

tents of the trunk, reveal a forger testing out his methods. The staining recipe of iron, manganese and chromium seems to have been Hinton's own, based on his knowledge of post-depositional processes affecting fossils in gravel.

Why the Piltdown gravels? Hinton was an expert on the geology of the Weald area of Sussex, in which Piltdown is located: Gardiner and Currant believe Hinton chose the Piltdown gravels precisely because they were entirely unfossiliferous, leaving Hinton a clear field to execute the fraud. Dawson and, through him, Smith Woodward, was led to the scene — and the rest is history.

Hinton knew Dawson was an incompetent geologist and would serve as the dupe; Dawson had already unknowingly traded a stone implement, stained to look old by Hinton, with Harry Morris, an expert on stone tools. This later turned up in Morris's collection labelled that it had been stained by Dawson with intent to defraud. Gardiner argues that Dawson is unlikely to have traded a flint he had faked with an expert such as Morris if he had done it himself.

The real victim seems to have been Smith Woodward, and the motive an argument about money. In 1910, Hinton wrote to Smith Woodward asking for vacation

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Hinton (centre) at work in the Natural History Museum.

employment cataloguing rodent remains at the museum. Woodward agreed, provided the payment of £130 was made after completion of the work, as was customary.

Hinton responded with a letter requesting that the sum be paid as a weekly wage, and detailing elaborate and costly plans for a catalogue. Woodward's reply, if any, does not survive, but as a senior figure (and experienced cataloguer) he is unlikely to have been impressed by the presumption of a junior colleague.

Whatever the outcome of the dispute, Hinton spent most of his subsequent career in the zoology department of the museum, away from Woodward's palaeontology department — even though much of Hinton's work concerned fossils.

Although the evidence of Hinton's responsibility appears strong, some doubts will inevitably remain among those who have studied the case closely. "It's a very convincing link between Piltdown and Hinton," says Chris Stringer, a palaeoanthropologist at the Natural History Museum. "But I still have my suspicions that Dawson was involved."

But Gardiner feels that the evidence for Hinton having been the sole hoaxer is now conclusive. He points out, for example, that Hinton was well-known for his elaborate practical jokes. The Piltdown fraud would have been an ideal way to get back at the pompous, stuffy keeper of palaeontology. Such suspicions are strengthened by the text of a letter Hinton wrote in 1954 to the evolutionary biologist Gavin de Beer — then the director of the British Museum (Natural History), now the Natural History Museum — after the fraud had been exposed.

"The temptation to invent such a discovery of an ape-like man associated with late Pliocene mammals in a Wealden gravel might well have proved irresistible to some unbalanced member of old Ben Harrison's circle at Ightham," wrote Hinton, a reference to his circle of Sussex-based geologist colleagues. "He [Harrison] and his friends [of whom Hinton was one] were always talking of the possibility of finding a late Pliocene deposit in the Weald." Given what we now know, this reads as almost a signed confession.

Henry Gee

Basic research wins out in UK spending

London. Policy changes introduced by the British government over the past ten years have resulted in a major shift in the balance of public spending on research and development (R&D), according to figures released by the government this week as part of its annual *Forward Look*.

In particular, basic research has increased its total share of such spending from 19.1 per cent in 1985–86 to 33.3 per cent in 1994–95. In contrast, government spending on 'experimental development' has fallen from 44.2 to 28.9 per cent of its total R&D expenditure, in line with its commitment to pass responsibility for such investment to the private sector.

Largely as a result of an accelerated shift in this latter direction, overall spending by the British government on civilian research and development, after a spurt at the beginning of the decade, has slowed down again. The figures published this week show that total expenditure on civilian R&D increased last year by only 1.4 per cent in cash terms — to a total of £3.17 billion (US \$4.79 million) — compared to a 4.4 per cent increase the previous year.

A major drop in spending on military research in particular has meant that the government's total R&D budget fell in cash terms for the first time for more than a decade. But spending on the science and engineering base — including the seven research councils — grew by 4.1 per cent, slightly higher than inflation. Basic research has increased its share of the civilian R&D budget from 35.1 per cent in 1985–86 to 54.9 per cent last year.

In terms of public spending on R&D by socio-economic objectives, the proportion spent on health has increased from 4.5 to 7.6 per cent between 1986–87 and 1994–95, and on environmental protection from 1.1 to 2.3 per cent. In contrast, spending on new energy sources has seen its share of the total fall from 4.4 to 1.1 per cent.

Sir Robert May, head of the Office of Science and Technology and the government's chief scientific adviser, says that a special effort has been made to make the report, which is intended to provide the government with an annual trans-departmental 'snapshot' of science spending, more readable than in the previous two years. □

'Unrealistic' misconduct plans under fire

Washington. A group of prominent US scientific organizations is urging the US Department of Health and Human Services (DHHS) not to implement many of the recommendations of an independent commission on scientific misconduct, describing them as "inadequate", "inappropriate" and "unrealistic".

In particular, the scientific bodies complain that procedures proposed by the commission would unduly favour the accuser, making an accused researcher "guilty until proven innocent". They also claim that the commission has rejected a succinct and precise definition of misconduct, replacing it with one that is that is "overly broad, legalistic and open-ended".

But the commission's proposals have been defended by its chairman, Kenneth Ryan, emeritus professor of obstetrics, gynaecology and reproductive biology at Harvard Medical School. Ryan claims that the critics have exaggerated the potential intrusiveness of the suggested measures and have little experience of the pressure that can be put on 'whistleblowers'.

The complaints, which came from a group of 50 professional societies organized by the Federation of American Societies for Experimental Biology (FASEB), have been made about a report of the Commission on Research Integrity (CRI), requested by Congress and released last November.

The comments are made in a press release and a letter sent last week by Ralph Bradshaw, the president of FASEB and a professor of biological chemistry at the University of California, Irvine, to William Raub, the head of a panel of DHHS officials set up to recommend how to implement the CRI report.

According to FASEB, the biomedical and bioscience organizations that endorsed the letter represent more than 285,000 scientists. The suggestions of the panel headed by Raub are due to be presented to Donna Shalala, the secretary of DHHS, within several weeks. She will decide if and how they are adopted. Raub said last week that his panel is "giving consideration" to the letter, but declined to comment further.

The 50 organizations, which include the American Federation for Clinical Research, the American Association of Immunologists and the American Society of Human Genetics, also argue that the commission's suggestions would impose "intrusive, expensive, and time-consuming" regulations on research institutions.

The CRI report was drawn up over 17 months by a 12-member panel of non-government scientists, ethicists and lawyers. It was mandated by Congress, in the National Institutes of Health Revitalization Act of 1993, to develop a new definition of research misconduct, to develop regulations

protecting whistleblowers and to suggest how to assure institutional compliance with DHHS regulations.

Current federal law defines scientific misconduct as "fabrication, falsification, plagiarism" or other practices that "seriously deviate" from those commonly accepted in the scientific community. Under the Public Health Service Act, this definition applies to all researchers funded by the National Institutes of Health (NIH), and some funded by other agencies.

In a report published in 1992, the National Academy of Sciences (NAS) recommended a narrower definition. It suggested that misconduct in science should be defined as "fabrication, falsification, or plagiarism, in proposing, performing, or reporting research" — excluding from the definition errors in data recording, selection, analysis and interpretation.

In contrast, the review commission suggests that the words "fabrication, falsification, or plagiarism" should be replaced by "misappropriation, interference, and misrepresentation". Some argue that this would

cover significant failures in data collection and storage; unlike the academy, for example, the CRI proposal would punish intentional omissions from a scientific paper if they falsified the entire report.

Ryan argues that the academy's definition of misconduct is too narrow. "They only want to call misconduct outright fraud. I think it's very hard to draw that fine distinction," he said. But the FASEB statement claims that the commission's proposed definition could "stifle" intellectual creativity by opening scientists to "unpredictable and ill-defined" charges of misconduct.

The FASEB statement also challenges the CRI's recommendations of protection for whistleblowers, saying that these would "make the accuser part of the investigating team" without holding them accountable for "violations of confidentiality, false statements, or other unlawful behaviour". But Ryan argues that the FASEB scientists, have been "shielded" from the bad behaviour of institutions towards whistleblowers, while the balance of power is often "heavily tilted toward the accused". **Meredith Wadman**

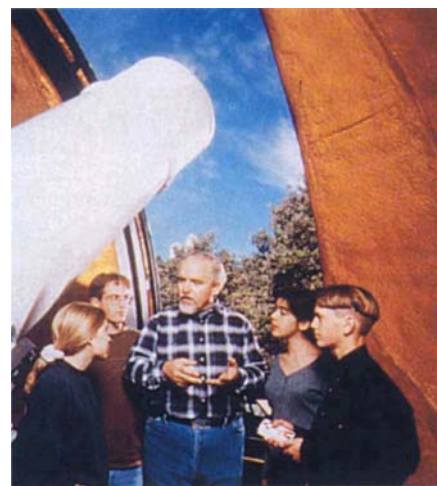
Remote stargazing draws the young

Geneva. Plans to extend a project that allows school pupils to operate research-grade telescopes at a distance received a boost last week, when its founder Gilbert Clark — an engineer at the NASA/Caltech Jet Propulsion Laboratory (JPL) in Pasadena, California — was selected from thousands of applicants as one of the five winners of the 1996 Rolex Awards for Enterprise, presented in Geneva last week.

The Telescopes in Education (TIE) project lets pupils equipped with a computer and modem operate Caltech's 24-inch reflector telescope at the Mount Wilson Observatory outside Los Angeles. Using a 'point and click' planetarium interface, they can steer the telescope to photograph galaxies, nebulae and other objects, and download the images.

The Caltech telescope, which was used during the Apollo programme to verify that the Moon was not covered in a thick dust layer that would have swallowed a landing spacecraft, was put into mothballs in the 1970s. Three years ago, Clark persuaded Caltech to let him reinstall the telescope at the Mount Wilson Observatory with the support of its director, Robert Jastrow, and NASA (the National Aeronautics and Space Administration).

With the help of volunteers, Clark refurbished the telescope, fitted it with charge-coupled-device (CCD) cameras, and modified it so that it could be operated at a distance by computer. Now the telescope is booked up months in advance by



Looking up: Clark with Caltech telescope.

schools. Richard Yessayan, a science teacher in Los Angeles, describes the impact of the project on children's interest in science as "overwhelming".

Two of the American schools taking part in the TIE programme have already recalculated the orbit of Pluto, producing data that will be used by NASA in planetary missions.

Encouraged by the general response, Clark says he now plans to extend the programme. He hopes soon to include another telescope at Mount Wilson, four in Australia, and a radiotelescope. "We need at least 20 for an international network," says Clark. **Declan Butler**

California to keep energy labs contracts

Washington. The US Department of Energy (DoE), in a decision that appears tinged with election-year politics, announced last week it will not put up for competitive bidding the contracts now held by the University of California (UC) for operating its three national laboratories — Lawrence Livermore, Lawrence Berkeley and Los Alamos — when the current agreements expire next year.

Senior DoE officials deny that political considerations have been involved in the decision. Nevertheless, the announcement includes a highly unusual congratulatory statement to the university from President Bill Clinton. And Charles Curtis, DoE's Deputy Secretary, mentioned Clinton's support for the university several times.

"Over the last five decades, the University of California made an enormous contribution to our success in winning the Cold War," Clinton said in his statement. "We look forward to working with [UC] to promote both our economic and national security."

Winning votes in California is crucial to Clinton's bid for re-election. "You're going

to see everything going to California that can go there" in the next several months, according to one Republican source, who points out that former President George Bush also steered contracts to the state during his failed 1992 re-election bid.

But Curtis insists the department had decided to stick with the University of California solely for non-political reasons, the most important of which, he says, is the need to maintain "continuity of management" at the two nuclear weapons laboratories — Los Alamos and Livermore — at a "sensitive time" in the post-Cold-War environment. "This is not a political issue by state," says Curtis. "This is not huge money to the University of California system; this is a service to the nation, as [UC] will emphasize."

The university has operated the laboratories since their inception, and receives about \$20 million a year to operate the three facilities. DoE said it would negotiate five-year extensions for all three facilities, which together cost \$2.5 billion a year to run.

According to Curtis, avoiding disruption at the weapons laboratories is especially

important in the light of DoE's need rapidly to put into place the elements of its "science-based stockpile management" programme for ensuring the reliability and safety of the nuclear arsenal without underground testing. These include the "accelerated strategic computing initiative" aimed at improving the modelling of nuclear weapons explosions, as well as new experimental instruments such as the National Ignition Facility, a \$1.1-billion laser fusion device to be built at Livermore.

While management of the third weapons laboratory, Sandia National Laboratories — now operated by the giant defence contractor Lockheed Martin Corporation — changed hands in 1993 with scarcely a ripple, Curtis maintained that a change at the UC laboratories would be much more disruptive. The research and development work at Sandia, which concerns the electronic and other non-nuclear components of nuclear weapons, "lends itself more to a conventional for-profit contractor," he says.

The university's management of Los Alamos came under fire last year from two Democratic members of New Mexico's congressional delegation for its lack of involvement with the local community. Representative Bill Richardson and Senator Jeff Bingaman had urged the DoE to solicit other bidders for the management contract, although Richardson later changed his mind.

Reflecting such concern, the extension of the Los Alamos contract will be conditional on a commitment by UC to tackle community development issues in northern New Mexico. The university will also be required to make unspecified management changes to improve operation of the laboratory's 'non-core' activities, such as environmental clean-up activities and safety and health operations.

David Kramer

Ministry merger prompts fears in Italy

Munich. Italy's ministry for universities and research, created only eight years ago to promote the needs of higher education and research more effectively, has been merged by the new centre-left coalition government with the much larger ministry of public education.

As part of the cabinet appointments announced last week, 57-year-old Luigi Berlinguer, a professor of law at the University of Siena, and head of the parliamentary group of the PDS, the former communist party which is now the largest party in the new parliament, will become minister for education and research.

Berlinguer has many years of experience in parliament, and is widely respected as a competent and efficient administrator. But the merger of the ministries has raised concern in the Italian scientific community that the needs of research could become submerged by the sheer weight of the new super-ministry.

Glauco Tocchini-Valentini, director of the CNR Institute of Cell Biology in Rome, argues that research "did not receive the necessary attention in the negotiations to form the new government". Similarly, Luciano Maiani, president of the INFN, the Italian nuclear physics research organization, told the daily paper *L'Unità* that the new ministry's task would be "very difficult", and he feared one of the two sides — either education or research — may suffer.

Romano Prodi, the new prime minister, who at one point was believed to favour the

former research minister, Antonio Ruberti, for the research post, has given no specific reason for his decision to combine the ministries, although it does help to reduce the total number of government departments.

A similar merger took place in Germany eighteen months ago, when the ministries of research and education were combined to form the grandly titled 'ministry of the future'. But the German scientific community still feels that this merger has reduced the political priority given to research.

Alison Abbott

Satellite quartet tunes up for the big night

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS
REASONS

ESA

London. The four European satellites (two are stacked on the right) that are due to be launched together next week by the European Space Agency's new Ariane-5 rocket to study the interactions between the solar wind and the near-Earth environment. The so-called Cluster quartet will gather information on the magnetic storms, electric currents and particle accelerations that occur in the space surrounding the Earth.

The satellites, seen here in Munich before being shipped to the launch site in Kourou in French Guyana, will also collect data on the ionized gases whose behaviour physicists are trying to reproduce in the laboratory to generate thermonuclear energy. Once in space, they will move into an eccentric polar trajectory around the Earth, which they will then follow in tetrahedral formation for the next two years. □

Universities protest against Australian spending cuts

Sydney. Australia's academic community is locked in a bitter conflict with the new conservative Coalition government over plans to cut operating grants to universities by up to 12 per cent, a reduction of up to A\$600 million (US\$470 million).

The universities are complaining that the new government is breaking pre-election promises to sustain basic funding for higher education, and to add extra funds for infrastructure and postgraduate scholarships (see *Nature* 381, 5; 1996).

They now face the prospect of having to introduce fees for undergraduate courses and of raising the deferred graduate tax, a unique Australian scheme for repaying part of the costs of higher education.

Last week, 37 vice-chancellors met Amanda Vanstone, the minister who heads the Department of Employment, Education, Training and Youth Affairs (DEETYA), and learned that their earlier fears of a cut of up to 10 per cent were too cautious.

The Treasury and Finance Departments, which are leading the government's efforts to achieve budget cuts of \$A4 billion in both 1996 and 1997, want Vanstone's department to save about \$A1.6 billion this year.

Slashing the department's employment and training activities, despite an unemployment rate of nearly 9 per cent, has not achieved the necessary savings, and the higher education sector, which received \$A4.6 billion a year in operating grants, has therefore been chosen for across-the-board reductions.

Universities use 30 per cent of these operating grants — supplemented by \$A400 million from competitive grant schemes — to finance their research activities, including the cost of infrastructure. They are upset by what they claim is an arbitrary approach to making cuts that appears to disregard long-term commitments and lacks any strategy from the new minister for higher education.

Vice-chancellors have united with staff, students and alumni in an unprecedented attack on the government. All universities are spelling out exactly what the effects of cuts will be. John Niland, vice-chancellor of the University of New South Wales, one of Australia's leading research universities, describes the cuts as hitting "like a cyclone", and says their impact would be "equivalent to closing down two of our largest faculties".

Another vice-chancellor claims that the cuts could lead to "a state of almost collapse". Results could include reduction in student numbers or increased student/staff ratios, as well as a freeze on a long-running claim by staff for increased salaries that now lag behind salaries in other countries.

They also warn that between 3,500 and

8,000 jobs could be lost, and that leading researchers may well choose to emigrate. The net effect would be to seriously damage Australia's reputation in Asia as an education and research provider earning around A\$2 billion a year.

Publicity given to the vice-chancellors' protests has forced Vanstone into her first public statement on the issue. But the former lawyer only hinted that universities "would have some flexibility in how they make savings".

Universities created from colleges of education and technology institutes in the reforms introduced in 1987 by the then Labor government, as well as smaller universities in regional centres, are most worried about their capacity to cope with cuts. These universities have been keen to achieve the level of research funding enjoyed by the leading eight universities.

The vice-chancellors are demanding to see the Prime Minister, John Howard, and the Treasurer, Peter Costello, who are blaming the former (Labor) Finance Minister and now Opposition Leader, Kim Beazley, for the alleged 'black hole' of A\$8 billion in government funds.

Vanstone has become a focus of opposition attacks in the Senate, where the government lacks a majority and is threatened with the delay or rejection of legislation on partial privatization of the publicly owned telecommunications company and on the reduction of the power of trade unions.

Frank Hambly, the executive director of the Australian vice-chancellors' committee, describes the government's policies as "a real threat, after we really believed we had iron-clad [funding] guarantees". He says that universities may have to respond by cutting courses, closing some faculties, reducing the campuses in multi-campus institutions or amalgamating some universities.

Meanwhile, Peter McGauran, the Science and Technology Minister, has signalled his desire for greater selectivity in research and development by speaking out in favour of an expanded space programme. In particular, he has announced a plan to launch small communications satellites with Russian-built rockets from Darwin, in the Northern Territory, near the Equator.

Other areas designated for priority funding, such as marine science, are expected to follow, at the expense of some current priorities, such as astronomy. McGauran's rejection of Australia's bid to join the European Southern Observatory (see *Nature* 381, 100–101; 1996) has generated widespread protests among astronomers. But so far their complaints have made little impact on the government.

Peter Pockley

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hands up: Gloria Steinem (left) is among the women supporting the global campaign.

US coalition counters breast gene patents

Washington. An international campaign to oppose the patenting of genes involved in susceptibility to breast cancer has been launched by Jeremy Rifkin, the long-time scourge of the US biotechnology industry. The campaign is being backed by a wide range of prominent women's rights activists and women's health and 'social justice' groups from both the developed and developing world.

In a statement due to be issued at a press conference in Washington this week, Rifkin describes efforts to patent and market commercially *BRCA1*, the first such gene to be discovered, as representing an "assault on women" that "denies them control over the most intimate aspect of their being, their bodies' genetic blueprint".

The statement, which has been endorsed by, among others, the author Betty Friedan, Gloria Steinem, the consulting editor of *Ms.* magazine, and Bella Abzug, co-chairperson of the Women's Environment and Development Organisation, also seeks support for legislation to protect genetic privacy and ensure non-discrimination by employers, insurance companies and others before the marketing of genetic screening tests.

Rifkin describes the campaign being launched this week as "the first genetic rights movement in history". He says that a worldwide coalition made up of more than 250 women's health and social justice groups plans to file a formal petition with the US Patent and Trademark Office challenging a patent application on the *BRCA1* gene and its mutations which has been filed by Myriad Genetics of Salt Lake City (see *Nature* 371, 271; 1994).

One of those who have endorsed Rifkin's statement is Vandana Shiva, president of the Research Foundation for Science, Technology and Natural Resource Policy in New Delhi, India. Shiva says that Indian women are worried that, once screening test results are available, "employers and insurance companies could be biased against women who have tested positive". □

Postgraduate funding should 'depend on research quality'

London. Only the most scientifically productive university departments should receive funds for postgraduate research from the Higher Education Funding Council for England (HEFCE), according to a review of postgraduate education in the United Kingdom published last week.

Such a strategy is necessary to ensure that leading universities maintain the standard of their research and continue to compete internationally, says the review panel, which was chaired by Martin Harris, vice-chancellor of the University of Manchester.

But many universities have reacted with concern to the panel's suggestion that available funds should go only to departments scoring Grade 3 or better (on an ascending scale of one to five) in the funding council's regular 'research assessment exercise'.

The Harris committee was set up to reconcile the increasing number of postgraduate students in the United Kingdom — now 21 per cent of the total student population, up from only 13 per cent in 1979 — with the limited funds now available.

The group was jointly commissioned to carry out the report by the HEFCE, the Standing Committee for Principals, and the Committee of Vice-Chancellors and Principals. Its recommendations will be presented to the HEFCE board later this month, and also to the government's National Committee of Inquiry into Higher Education, chaired by Sir Ron Dearing.

According to Harris, the committee

decided not to recommend limiting student numbers to make do with currently available funds — as was done with the undergraduate sector — on the grounds that this would have restricted postgraduate courses. But the report also opposes the present practice of using resources from the 'already inadequate' undergraduate funding to accommodate the growth in the postgraduate sector.

In general, it endorses a market-led system for planning postgraduate education, leaving institutions free to expand provision beyond funded numbers either by charging additional fees or by making efficiency savings. Institutions would have to charge their students the real cost of courses.

The review group admits that a market-based system could mean that only the more fashionable areas of research might attract adequate numbers of postgraduate students, and that interdisciplinary study could suffer as a result. To compensate for this, it urges the research councils and the British Academy to ensure that proper attention is paid to all areas of research.

The group also recommends a code of practice for institutions to ensure that they offer appropriate facilities and supervisory arrangements for postgraduate student research. To clear up confusion about the definition of courses, it suggests a nationally applied set of criteria. Accurate descriptions of courses could eventually become a condition of government funding. **Anju Sharma**

Japan needs centre of excellence for brain research, council says

Tokyo. A proposal to set up a centre of excellence in brain research has been endorsed by the general assembly of the Science Council of Japan (JSC), a government advisory body attached to the Prime Minister's Office. A joint-use research facility and information centre, increased funding and greater cooperation between funding agencies are also recommended for the discipline.

The JSC is an elected body of scientists that advises the government on matters related to science and technology. The organization's head, Masao Ito, is one of Japan's leading brain scientists.

The proposals have been put forward in a report drawn up by a special committee on brain research and the mind. The report points out that although four different Japanese government ministries and agencies spend about ¥6 billion yen (US\$56 million) a year on brain research, this is an order of magnitude less than the funding US brain researchers receive from the National Institutes of Health alone.

The report says that funding should therefore be increased and distributed more efficiently by encouraging closer cooperation between funding agencies. According to Miwatani Toshio, a professor at Okayama Prefectural University and a member of the JSC Special Committee, the present system, with no central body to coordinate funding, leads to an overlap in research activities.

The proposed centre of excellence should be established at one of Japan's existing brain research laboratories, says the report. It would form the hub of a research network linking not only researchers in the physical sciences but also social scientists and others able to improve the understanding of the human mind.

Most ambitiously, the committee also suggests that a new cooperative-use research and information centre be set up to provide researchers in the field with advanced equipment and other facilities that individual institutions cannot afford. But Miwatani admits that funding restrictions mean that this recommendation will be difficult to implement.

Michio Oishi, director of the Ministry of International Trade and Industry's Institute for Bioscience and Human Technology, says that the government is under no obligation to take the council's advice. But he also points out that last year's science and technology basic law (see *Nature* 378, 227; 1995) has strengthened the scientific community's hand in its dealings with the Ministry of Finance, thus improving the prospects that the JSC's recommendations will be implemented. **Stephen Barker**

Bidding heats up for protein database

Paris. Prospects brightened this week for the rescue of the important SWISS-PROT protein database, threatened with closure next month following a decision by the Swiss government to stop funding for the database. Both the Swiss authorities and the European Commission are reported to be seeking solutions to the situation, after Amos Bairoch, the database's founder, received more than 1,000 letters of support from researchers worldwide. Switzerland has suddenly discovered that "after watches and chocolates, it is best known abroad for SWISS-PROT", says Bairoch.

Meanwhile, the US biotechnology company Incyte has offered to buy the database. But SWISS-PROT is considering this as a "last resort", and would prefer the database to remain in the public domain. A sale would also be likely to lead to legal conflict between SWISS-PROT and the European Bioinformatics Institute in Cambridge, England, which is

a major partner in the database and has a policy of making data freely available.

The publishing company, Reed-Elsevier — which has extensive database activities — is also reported to have put in a bid for the database. SWISS-PROT says this would be more acceptable than selling to a drug company, but would be unpopular with users, as Reed-Elsevier would inevitably charge for access. Many drug companies are also loathe to see the database fall into any one company's hands, and around thirty of them — including Glaxo-Wellcome, Hoffmann-La Roche, SmithKline Beecham, and Millennium — are said to have offered donations to keep the database afloat.

The same companies are also said to have offered to set up a consortium to fund the database. Bairoch says that even if public funds become available, ensuring the database's long-term survival may require charging commercial users for the first time. **Declan Butler**

Indian strategy takes N-test talks to brink

Paris. Prospects of a treaty banning all nuclear tests were hanging in the balance this week, as what may be the final round of negotiations began in Geneva. Although consensus has been reached on many aspects of the text, it is still possible that the entire treaty could be derailed by India, which feels that the text does not make an adequate commitment to nuclear disarmament.

The current draft of the Comprehensive Test Ban Treaty (CTBT) goes further than its strongest advocates could have hoped for this time last year. Negotiators from about 60 countries meeting at the Conference on Disarmament in Geneva are keen to reach agreement by next month, when presidential elections will be held in Russia.

This would allow the text to be sent to the UN General Assembly in September, and to take effect by next year. If a treaty is not agreed by next month, any new Russian government may demand further changes, causing inevitable delays.

"A draft with a thousand sections still in square brackets [that is, still to be agreed] will get swept away" by major political upheavals, says Christopher Paine of the Natural Resources Defense Council in Washington, arguing that an agreed treaty will be better able to resist political change.

Broad resolve to bring the negotiations to a swift conclusion was apparent in Geneva last week. In particular, China dropped its previous opposition to the June deadline for reaching agreement on the text. In a major shift, it also hinted that it may drop its insistence that the test ban treaty should permit nuclear explosions for civilian purposes. Observers say China may be ready to agree to a compromise that would outlaw such explosions, pending a review in 10 years' time. This face-saving device is tantamount to a ban, says Paine.

Russia last week helped to achieve a speedy conclusion of the negotiations by formally agreeing to a ban on all nuclear tests, however small. This brings it into line with the other four nuclear weapons states — the United States, China, France and the United Kingdom — which recently agreed to accept a so-called zero-yield treaty (see *Nature* 376, 540; 1995).

In the past, the nuclear weapons states had pressed to keep open a loophole in the treaty allowing them to carry out low-yield nuclear explosions of up to 1 kiloton (see *Nature* 376, 283; 1995). But non-nuclear-weapons states have opposed this, arguing that it would eliminate disarmament from the goals of the CTBT by allowing nuclear weapons states to continue modernizing their arsenals.

But the disarmament goals of the treaty remain the major sticking point. India in particular is concerned that although the

treaty would encourage non-proliferation, it would do little to prevent the five nuclear weapons states from keeping their weapons.

India feels that its ambitions to become a regional power depend either on becoming a nuclear weapons state or on shifting the global balance of power in its favour by persuading the existing ones to disarm. It therefore refuses to sign a treaty that lacks firm assurances from the five nuclear weapons states that they will eliminate their nuclear arsenals.

In particular, India wants the treaty to include a ban on any "qualitative" improvements in nuclear weapons, in order to prevent the nuclear weapons states from modernizing their arsenals using the computer and experimental simulation techniques to which they alone have access. (Nuclear powers argue that these techniques are needed to maintain the safety of their weapons stockpiles, see *Nature* 380, 8; 1996.) Similarly, India wants the treaty to specify a goal of achieving nuclear disarmament within ten years.

But many non-nuclear weapons states are concerned about India's strategy. Although sympathetic to India's complaints, they acknowledge that the nuclear weapons

tion techniques to offset the effects of a ban. The United States, for example, is proceeding with a large programme of subcritical testing this year at its underground facilities at Nevada.

"They are giving the impression that they want to get away with as much as they can [within the treaty]," says one observer, who describes the US decision to proceed with underground subcritical testing as a "red rag to India".

One compromise would be to include India's demands in general terms within the preamble to the treaty but not within the treaty itself. Most states are expected to try to persuade India to accept this formula, and India may eventually drop its demands, having succeeded in emphasizing that the treaty is about disarmament as well as non-proliferation.

But another hypothesis is that India will use the inevitable refusal of the nuclear weapons states to accept its demands as an excuse for refusing to sign the agreement at all. Such speculation has intensified since the Hindu nationalist Bharatiya Janata Party won the most seats in the recent election, as its manifesto called for the deployment of nuclear weapons.

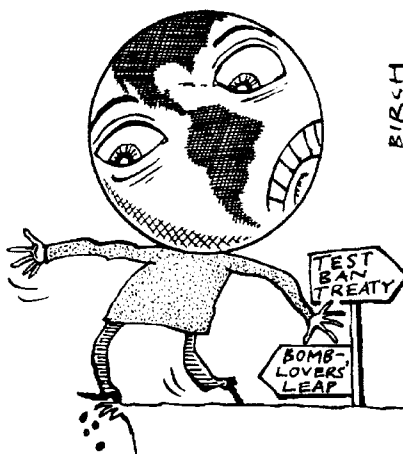
Indeed, many predict that India is poised to declare itself a nuclear weapons state. This could cause China and Pakistan to withdraw from the negotiations as well, leaving hopes of a treaty in tatters.

It remains unclear whether India will sign the treaty at all, whether the treaty could be agreed without it, and what value such a treaty would have. The United Kingdom has proposed that the treaty should take effect only if it is ratified by at least the five nuclear weapons states and the three threshold weapons states. Pakistan has since endorsed this proposal, and China and Russia are said to be coming around to this position.

But many other countries, including the United States, are said to be furious about this proposal, arguing that it would virtually legitimize the status of India, Pakistan and Israel as nuclear weapons states, while allowing any of the three countries to hold the entire treaty hostage. "Entry into force is turning into a central fight" in the negotiations, says Rebecca Johnston, from *Disarmament Intelligence Review*.

Others want the treaty's entry into force to depend on its ratification by all 68 countries that have nuclear fuel cycles. But, as India may refuse to sign the treaty, others are keen to get the best agreement possible by next month, and to postpone the decision on who signs and when. Such tactics have been used in the past to overcome similar obstacles to the ratification of both the chemical weapons convention and the nuclear non-proliferation treaty.

Declan Butler



states will not accept its demands, and feel that by pursuing them India could wreck the entire treaty. They advocate the more pragmatic approach of getting a test ban treaty first, and afterwards working further towards disarmament.

A ban on testing would prevent the five nuclear weapons states from developing more sophisticated arsenals, while freezing programmes in both non-nuclear-weapons states and the undeclared nuclear weapons states, India, Pakistan and Israel. By halting the arms race, supporters of this strategy argue, a test-ban treaty would set the stage for further rounds of weapons cuts.

At the same time, the nuclear weapons states may be undermining their attempts to convince others of the validity of this approach by investing massively in simula-

Republicans accused of 'risky experiment' in cutting budget

Washington. Republicans in the US Congress were last week accused by a senior official of the National Science Foundation (NSF) of carrying out "a very risky experiment" by curbing science and technology funding in an attempt to eliminate the budget deficit by the year 2002.

"The science and technology budgets, and the education related to [both], are extremely important to the nation," said Anne Petersen, who is both the deputy director of the \$3.2-billion agency and its chief operating officer. At a time when scientific opportunity has "never been greater", Republican plans to eliminate the \$144-billion deficit will "severely impair" the government's ability to make good on the promise of science, Petersen said in Washington.

On the day she was speaking, the US House of Representatives split almost entirely along party lines when it voted by 226 to 195 to pass a budget resolution for 1997 based on a gradual reduction in spending on various civilian science agencies — including for example, the NSF and the National Aeronautics and Space Administration — from \$16.5 billion in 1997 to \$15.6 billion in 2002. □

Chernobyl academics in jail

London. Two academics from Belarus, jailed in Minsk for organizing a rally to commemorate the tenth anniversary last month of the Chernobyl nuclear accident, have been on hunger strike for nearly a month. Yuri Khadyka, a 56-year-old physicist with the Belarus Institute of Physics, and Viachaslaw Siwchyk, a 33-year-old geologist, were arrested on 26 April. They are described as in "critically poor health, severely weakened and disoriented" after almost three weeks without food. Relatives are not allowed to bring them medication.

Approximately 400 Belarus academics have added their names to a petition calling for the immediate release of Khadyka and Siwchyk. As many as 50,000 people are believed to have taken part in the rally, which was organized with the permission of the Minsk authorities. The police were later ordered to break up the demonstration and more than 200 people were arrested. □

EU cuts radiation dose limits

London. The Council of Ministers of the European Union (EU) has ordered an 80 per cent reduction in the maximum quantity of ionizing radiation to which a member of the public can be exposed. Under a revised directive issued earlier this month, a member of the public in an EU member state can be exposed to a maximum of 1 millisievert (mSv) each year, compared to the present limit of 5mSv.

The directive also revises exposure limits for radiation industry personnel to 100 mSv in a consecutive five-year period, an average limit of 20mSv per year. This amounts to 40 per cent of the current 50mSv yearly limit. EU member states have been given four years to implement the new directive through domestic legislation. □

Funding for 'proteome' project

Sydney. The pharmaceutical company Glaxo Wellcome is due to finalize a A\$1.9-million (US\$1.46-million) contract today (23 May) to fund 'proteome' research at the University of Sydney in New South Wales, Australia. A proteome is the total amount of protein in any living cell that is encoded by each cell's DNA. Proteome research is coming to be recognized as the 'flipside' of the Human Genome Project.

The funding will allow Ian Humphery-Smith, a microbiologist at the university, to look at the feasibility of establishing a large-scale facility for proteome analysis. The work will be carried out at Humphery-Smith's new laboratory in the National Innovation



STREAMLINE
rProtein A
the most attractive
adsorbent for
purification.

No wonder, it's

Planning of US AIDS research

SIR — Your News report headed “AIDS office opposes outside planning” (*Nature* **381**, 5; 1996) draws a misleading interpretation from our comments to Congressman John Porter during testimony before the US House of Representatives appropriations subcommittee on health, education and labour.

Far from opposing the role of the outside scientific community in planning AIDS research activities, the National Institutes of Health (NIH) and the Office of AIDS Research (OAR) in particular have strongly encouraged such advice. The statement that “the direction of AIDS research should remain firmly in the hands of institute directors and their advisers” is a misinterpretation of the view we expressed, which was that peer review should remain a process devoted to the determination of the scientific merits of a grant proposal and separate from the planning process. Indeed, the inclusion of nongovernment experts from academic institutions, industry and the AIDS community in the planning and evaluation processes has been a hallmark of the OAR.

The NIH is currently preparing an implementation plan based on the recommendations from the National AIDS Research Program Evaluation Working Group led by Arnold Levine.

Harold Varmus

(Director)

William E. Paul

(Director, Office of AIDS Research)

National Institutes of Health,

Bethesda, Maryland 20892, USA

IQ and censorship

SIR — We welcome the fact that Christopher Brand's book *The g Factor: General Intelligence and its Applications*, has been reviewed in *Nature*¹, and we regret its withdrawal by the publishers, John Wiley & Sons. As departmental colleagues of the author we should like to make the following points.

(1) Without commenting on the content of this book, we feel that it deserved publication through the normal channels, not suppression. In an eloquent argument which could apply equally to Brand's book, James Flynn² made the case that J. P. Rushton's views — with which Flynn disagreed absolutely — deserved publication and then attempted scientific refutation rather than suppression. Flynn argued that “the truth can never be racist, nor can telling the truth as you see it, assuming

there is no evidence of wilful neglect of evidence”. In summary, Wiley has threatened academic freedom.

(2) The sensational treatment of this affair in the non-scientific press threatens to taint sensible research on intellectual abilities, as Brand may in future be used as a convenient bogeyman to discredit work in a broad and important area. (As a parallel, Sir Cyril Burt's name is used to give a foul smell to research quite unconnected with his own; for instance, by Kitcher in his assessment of ‘pop’ and human socio-biology³.)

(3) The data presented in Brand's book may usefully be read alongside the authoritative views recently endorsed unanimously by an American Psychological Association (APA) task force⁴. From Brand's book, your reviewer, N. J. Mackintosh, singles out two details for comment: the *broad* heritability of IQ, and the correlation between IQ and “inspection time” (IT). In the first case, the APA consensus estimate of adult *broad* heritability agrees with Brand rather than Mackintosh. On the second, less controversial issue, the APA experts are with Mackintosh rather than Brand. As this latter disagreement is not about observed values, but rather about how to extrapolate from these to the ‘true’ IQ–IT correlation in the ‘normal population’, this example serves to make the point that determination of truth requires discussion. Withdrawal of the book will not advance our understanding of the truth in the cases where Brand's interpretation of evidence can be disputed.

P. G. Caryl

I. J. Deary

Department of Psychology,

University of Edinburgh,

7 George Square,

Edinburgh EH8 9JZ, UK

1. Mackintosh, N. J. *Nature* **381**, 33 (1996).

2. Flynn, J. R. *Psychologist* **2**, 363–366 (1989).

3. Kitcher, P. *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (MIT Press, Cambridge, Massachusetts, 1985).

4. Neisser, U. et al. *Am. Psychologist* **51**, 77–101 (1996).

Global surface temperatures

SIR — Gordon *et al.*¹ comment, somewhat misleadingly, on the press release issued by the Met Office (jointly for the University of East Anglia) on 1995 global temperatures. We should like to make the following comments.

The December 1995 global surface temperature anomaly was a combination of sea surface temperatures (two-thirds of the globe) and land (near-surface air) temperature, the latter calculated using a strong relationship² with (final, not provisional) 500-hPa geopotential fields. This is

rather more than an “educated guess”. The final estimate of the global annual anomaly remained the same as that in the press release. We pointed out that “differences of a few hundredths of a degree between global average temperatures in individual years are not significant”.

The temperature anomaly for the globe was not “determined largely by that for the Northern Hemisphere”. The global mean is a straightforward average, with cosine (latitude) weighting, of anomalies for each 5° × 5° square where sufficient data exist. If, as is sometimes the case, there are more data-sparse squares in the Southern Hemisphere (SH), then a slight Northern Hemisphere (NH) bias can occur. In fact, in 1995, the NH temperature anomaly was 0.55 °C and the SH anomaly was 0.23 °C, giving a global anomaly of 0.39 °C, not significantly different from the 0.40 °C in the press release. We said in the press release that the SH was “relatively less warm”; nevertheless, our record shows it to be (along with 1990, 1991 and 1993) the third warmest since 1860.

The press release clearly referred to the “surface temperature of the earth”, that is, the region that we inhabit. The MSU satellite instrument averages temperatures over a deep layer from the surface through most of the troposphere; it need not agree closely with surface temperature in its decadal trends or shorter-term fluctuations. The relationship between the two is discussed extensively in the forthcoming 1995 report of the Intergovernmental Panel on Climate Change³ and is the subject of continuing work at the Hadley Centre, the University of Alabama^{4,5}, the National Aeronautics and Space Administration⁶ and elsewhere.

David E. Parker

Hadley Centre for Climate
Prediction and Research,
Meteorological Office,
Bracknell, RG12 2SY, UK

Philip D. Jones

Climatic Research Unit,
University of East Anglia,
Norwich, NR4 7TJ, UK

1. Gordon, A. H., Bye, A. T. & Byron-Scott, R. A. D. *Nature* **380**, 478 (1996).

2. Wallace, J. M., Zhang, Y. & Bajuk, L. J. *Climate* **9**, 249–259 (1996).

3. IPCC Working Group 1. *Climate Change 1995 — The Science of Climate Change*. Second Assessment Report of the Intergovernmental Panel on Climate Change (eds Houghton, J. T. et al.) (Cambridge Univ. Press, in the press).

4. Christy, J. R. & Spencer, R. W. *Clim. Change* **30**, 97–102 (1995).

5. Christy, J. R. *Clim. Change* **31**, 455–474 (1995).

6. Hansen, J. et al. *Clim. Change* **30**, 103–117 (1995).

corres@nature.com

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only, or e-mailed (with full mailing address) to corres@nature.com

Morgagni and the impact factor

Women scientists

SIR — The recent debate about the impact factor (IF) and citation index as a measure of scientific value, especially in competition for university posts in Italy, is a reminder that this is nothing new. The University of Padua's medical faculty recently introduced an 'impact factor' threshold for future members of its scientific and development commissions¹. This procedure has a long but forgotten tradition.

Throughout his long and prolific career, Giovanni Battista Morgagni (1681–1771), probably the most renowned anatomist and pathologist of the eighteenth century, and a professor at Padua from 1712 to his death, collected in a register all citations of his work. This was only one aspect of Morgagni's self-promotion, which was also aimed at attracting the attention of other scholars so that they would disseminate information about him.

During his lifetime, Morgagni was honoured with a monument commissioned by the Natio Germanica students who attended Padua's Studio. During the nineteenth century, a century of great change in the field of medicine, Morgagni's supremacy in introducing the 'anatomic idea' into pathology was not forgotten, as Rudolph Virchow demonstrated². Even today, Morgagni is the subject of a plentiful literature and is often cited in databanks. To test the contemporary 'impact' of Morgagni's work would be difficult; fortunately, Morgagni himself has done a good part of this work, registering for more than 60 years (from 1704 to 1768) all the citations of his name³. Morgagni was perfectly aware of the importance of controlling the circulation of his work and its acceptance by colleagues. It is noteworthy that much of his collection of private letters — often intended from the first for publication — was concerned to promote his image as a man, physician and scientist. Moreover, an important part of his work *Adversaria Anatomica* was devoted to refuting the criticisms of his detractors.

Morgagni did not limit himself to collecting citations about himself, but also promoted them by keeping his colleagues informed about the progress of his research. As soon as accounts of his work were published, he selectively distributed them among a few authorities and then made their favourable assessments known. His autobiographies are rich in detailed references to the citations⁴.

How much did Morgagni's self-promotion contribute to his fame among his contemporaries? A comparison could be made with his friend and colleague, Antonio Vallisneri (1661–1730), a professor at Padua from 1700 to his death. Vallisneri was no less active in his own field of natural history than the anatomist⁵, yet much of Vallisneri's innovative research was ignored by

his contemporaries and by most historians of biological thought.

Citation analysis does not always indicate the quality and impact of an individual's work. The best ratio between production and promotion in science is perhaps the critical point of the scientific quality valuation. Using citation analysis and IF *tout court* as a mechanical replacement for careful human judgement cannot change a John Doe into Morgagni.

Maurizio Rippa Bonati

*Istituto di Storia della
Medicina, Università,
via G. Falloppio 50,
I-35121 Padova, Italy*

Andrea G. Drusini

*Department of Biology,
University of Padova,
Via Trieste 75,
I-35121 Padova, Italy*

1. Calzà, L. & Garbisa S. *Nature* **374**, 492 (1995).
2. Virchow, R. *Morgagni und der anatomische Gedanke* (Verlag von August Hirschwald, Berlin, 1894).
3. Morgagni, G. B. *Catalogo cronologico (1704–1768) delle opere a stampa nelle quali sono citati gli scritti di G.B. Morgagni, o si parla di lui e del suo merito scientifico*, Vol. 16. Ms. cc. 95 (Biblioteca Comunale di Forlì, Fondo G. B. Morgagni).
4. Morgagni, G. B. in *Opera Postuma*, Ms. Laurenziano Fondo Ashburnhamiano (ed., Pazzini, A.) Vol. 1, 227–159. (Istituto di Storia della Medicina dell'Università, Roma, 1964).
5. Omodeo, P. *Antonio Vallisneri fra Padova e l'Europa scientifica del primo settecento*, ANMS 9(1), 3–4 (1982).

Not revisionist

SIR — In his review (*Nature* **380**, 399–400; 1996), Paul Weindling calls my recent book *Biologists under Hitler* a "revisionist verdict" that denies the impact of the Jewish emigration on modern biology in post-war Germany. Indeed, I stated that, despite serious losses, "emigration alone is not a sufficient explanation for the lag in biology in Germany after World War II". But I attributed this lag to the emigration *and* to the moral failure of virtually all German non-Jewish biologists. I said that they violated the scientific principle of universalism by accepting the expulsion of their Jewish colleagues almost unopposedly, collaborated loyally with the Nazi government, and thus broke off dialogue with their international colleagues. As a consequence, German biologists — with the exception of a very few individual scientists — did not participate in international scientific exchange for several years after 1945. I pointed out that this self-inflicted isolation of German biologists contributed decisively — in addition to the forced emigration — to the late onset of molecular biology in post-war Germany. Should this reasoning be called a revisionist verdict?

Ute Deichmann

*Institut für Genetik,
Universität zu Köln,
Weyertal 121, D-50931 Köln, Germany*

SIR — I read June Goodfield's review (*Nature* **380**, 306–307; 1996) of Margaret W. Rossiter's new book, *Women Scientists in America: Before Affirmative Action 1940–1972*, with great interest, and I plan to buy the book when I go "back to where I came from". Where I came from most recently is the Massachusetts Institute of Technology (MIT). At MIT there are four women faculty in a biology department of 60 faculty, and one of those women asked me once, "Where are the women?" I don't know, but I know that we are out of science before we reach faculty level. If we remain, we drift about from postdoc to postdoc, and, in my case, from country to country.

But the reason for this letter is to defend my own women's college, Smith. I loved MIT and the informal, creative intelligence of its biology professors, but nowhere has ever been as intellectually fulfilling for me as my women's college, and I certainly do not feel "betrayed". At Smith, I was surrounded by people who not only believed I could excel in mathematics, physics and biology, but who also *expected* it. Since I recently applied for a directorship of a programme on women and science, I have been thinking about why I am one of four women at my career stage in the Max-Planck-Institut für Molekulare Pflanzenphysiologie, and why we, as postdocs, are the most advanced women at this institute. In my case, the reason is in great part Smith College. Without the four years of being trained by Jill Ker Conway to believe very firmly in our abilities, I doubt whether I would even have a PhD, much less have had postdoctoral fellowships at Harvard, MIT and the Max-Planck and my own independent research project.

Goodfield refers to "...a cadre of highly qualified women who had no place to go except back to where they came from — the women's colleges". Exactly! Highly qualified women were and are coming from women's colleges. More than 75 per cent of the graduates from Smith College go on to graduate work. Without education in women's colleges, there wouldn't even have been a question of where women were going because there wouldn't have been any highly educated women. Where did the women's movement start to get any of us here in science in the first place? It certainly did not start at the Harvard Faculty Club. It was with Gloria Steinem, graduate of Smith College, founded by a woman named Sophia Smith who fought against the notion prevalent in her time that higher education was bad for women's health. Women's colleges put us in science and they will foster the women who come up with the ideas to keep us in science.

Ellen V. Kearns

*Max-Planck-Institut für
Molekulare Pflanzenphysiologie,
Karl-Liebknecht-Str. 25, Haus 20,
14476 Golm, Germany*

Errors in protein structures

SIR — In 1990 Brandén and Jones¹ wrote in a Commentary to *Nature*: “Protein crystallography is an exacting trade, and the result may contain errors that are difficult to identify. It is the crystallographers’ responsibility to make sure that incorrect protein structures do not reach the literature.” Your recent leading article² about obligatory deposition of macromolecular coordinates and the underlying experimental data³ revived this debate.

We have recently carried out a search for anomalies in 3,442 structures from the Brookhaven Protein Data Bank using our program WHAT_CHECK (which can be retrieved by anonymous FTP from swift.embl-heidelberg.de or from pdb.pdb.bnl.gov). We have analysed the deposited coordinates without using the underlying experimental data. We assume that the standard deviations in parameters such as bond lengths and bond angles that are observed in small molecules can be used as a standard of truth⁴.

Our analyses reveal a much larger number of 4σ deviations than one would expect. Some of these outliers will reveal a novel feature (for example, a property typical for transmembrane proteins, of which there are now only a few in the database). Another fraction represents cases where the experimental data simply cannot be described by ‘normal’ coordinates (for example, two alternative conformations for a side chain, but insufficient data to resolve them). However, we also detected many straightforward errors. These range from wrong atom names or bond angles that are more than 10° off, via threonines with wrong C- β chirality and tryptophans with a 90° angle between the two rings, to major errors such as proteins that are misthreaded or solved in the wrong spacegroup (see table).

Many of these problems are not crucial to the casual browser. But for purposes of detailed biological studies or large surveys, it is important that these irregularities do not go undetected. The Protein Data Bank is aware of the importance of structure verification: it is increasingly making validation tools part of its production process, exchanging experience and software with European colleagues working in this field. (The BIOTECH structure validation server provides structure verification tools via the World Wide Web. WHAT_CHECK is part of this server. WWW addresses are: <http://biotech.embl-heidelberg.de:8400/>,

<http://biotech.embl-ebl.ac.uk:8400/> and <http://biotech.pdb.bnl.gov:8400/>.)

The current verification software can

A FEW OF THE ERRORS IN THE LITERATURE...

Inconsistent symmetry information	19 files
Transformation matrix has determinant not equal to 1.0	5 cases
D amino acid	183 cases
Atom too close to symmetry axis leading to a clash	332 cases
Structure probably solved in wrong space group	24 files
Much too high Matthews’ coefficient ($V_m > 7.0$)	69 files
B-factors over-refined	533 files
Cell dimension off by more than 0.5%	1,914 files
Atomic occupancies negative	
or larger than 1.0	43,934 cases
Bond length deviates more than 4σ	61,051 cases
Bond angle deviates more than 4σ	309,186 cases
Atoms more than 0.4 Å too close	
to each other	265,290 cases
Side chain of His, Asn of Gln needs 180° flip	19,906 cases

The 1,159,804 outliers in Protein Data Bank data sets reflect discrepancies with conventions, statistical outliers and probable errors. Of the 76 classes of problems only 13 are listed in this table. The complete tables, full reports about every entry that we tested and detailed descriptions of all tests are available from <http://www.sander.embl-heidelberg.de/pdbreport/>

already solve some of the detected problems automatically. Others require manual electron density inspection. We hope that structure factor deposition will soon be as much an accepted practice as coordinate deposition, so that in the future more problems can be detected and more of the million or so found so far can be solved.

Rob W. W. Hoofft

Gert Vriend

EMBL,

Meyerhofstrasse 1,

D-69117 Heidelberg, Germany

Chris Sander

EMBL-EBI,

Hinxton Hall, Hinxton,

Cambridge CB10 1RQ, UK

Enrique E. Abola

Protein Data Bank,

Brookhaven National Laboratory,

Biology Department, Building 463,

Upton, New York 11973, USA

1. Brandén, C. -I. & Jones, T. A. *Nature* **343**, 687–689 (1990).

2. *Nature* **379**, 191 (1996).

3. Baker, E. N. et al. *Nature* **379**, 202 (1996).

4. Engh, R. & Huber, R. *Acta crystallogr.* **A47**, 392–400 (1991).

Paradoxical error

SIR — Last year I published in *News and Views*, under the title “Multiple Kauzmann paradoxes” (*Nature* **373**, 475–476; 1995), an account of a paper, somewhat controversial as it proved subsequently, by two chemists, K. Kishore and H. K. Shobha (*J. chem. Phys.* **101**, 7037–7047; 1994). The paper discussed the extension of the well-known Kauzmann paradox, referring to an extrapolated temperature domain in which the entropy of a supercooled liquid is below that of the cor-

responding crystalline structure, to the relationship between a supercooled vapour and the corresponding liquid phase. Kishore and Shobha claimed to show that there is an upper limiting temperature above which a superheated liquid would indeed have a higher entropy than its vapour (‘entropy crossing’).

P. G. Debenedetti, M. M. Atakan and R. J. Speedy have now published (*J. chem. Phys.* **104**, 5349–5350; 1996) a rebuttal of Kishore and Shobha’s treatment in terms of spinodal theory in the vicinity of the critical state, taking into account pressure, which Kishore and Shobha failed to do. They demonstrate that “the entropy of a superheated liquid and a vapor at the same temperature and pressure are never equal except at the critical point, where both phases are identical”, and show how Kishore and Shobha arrived at what now appears to be their erroneous conclusion.

Robert W. Cahn

Department of Materials

Science and Metallurgy,

University of Cambridge,

Cambridge CB2 3QZ, UK

The last word

SIR — Webster and Erickson (*Nature* **380**, 386; 1996) call for a word to “designate the last person, animal, or other species in his/her/its lineage”. They give several possibilities, preferring ‘endling’.

In supporting instead ‘ender’, my colleague Ralph Elliott, a lexicographer in this university, points out that the *Oxford English Dictionary* attributes the word first to Chaucer and gives as one of its meanings ‘He or that which puts an end or termination to anything’. The word seems not to be in use for any other purpose.

David Craig

Research School of Chemistry,

Australian National University,

Canberra 0200, Australia

SIR — ‘Endling’ has a somewhat pathetic feel to it, similar to ‘foundling’. The suffix -arch means ‘leader’ and is used in the words matriarch and patriarch, meaning one who rules a family, clan or tribe. Terminal means ‘forming an end or boundary’.

I suggest ‘terminarch’ to designate the last of lineage. It has a much stronger and more positive ring.

Elaine Andrews

RPR Gencell,

5301 Patrick Henry Drive,

Santa Clara, California 95054-1114, USA

SIR — The last remaining is a relict — a word with a decent Latin root, still used and understood.

There is no need invent a new one.

Mark Smith

Ballacurn, Ballaugh,

Isle of Man IM7 5EU, UK

A New Deal for national science policy

Steve Fuller

What do rank-and-file scientists think the aim of peacetime national science policy should be? Should it be economic competitiveness? That is not what they said the last time they were asked.

A 'NEW DEAL' is needed for national science policy in peacetime. Most would agree that research priorities, funding levels and accountability mechanisms have been distorted in the past by concern about national security. But what would it mean to remove these distortions? The last time this question was asked was after the Second World War, during the debate over the establishment of the US National Science Foundation (NSF). But because the NSF ended up incorporating many wartime science policy mechanisms into its day-to-day practices, we shall have to explore the road not taken, the one that would have led back to the original New Deal, President Franklin Roosevelt's policy of economic reconstruction in the 1930s.

Even before the end of the war, the heads of many US state universities had begun to worry that peer-reviewed funding procedures in peacetime would exaggerate differences in resources between elite and non-élite institutions that had arisen from selective corporate investment in research and development (R&D) before the war¹. Peer review, long a mechanism for evaluating the suitability of finished research for publication, was extended during the war to assess a scientist's worthiness for doing research in the first place. Jointly threatened by the prospect that the Nazis might devise the first nuclear weapon, the Allied physicists put aside their theoretical differences and focused on bringing together the colleagues most likely to help them quickly achieve their goal. There was little disagreement about whom to select, and the job got done as planned.

Old prejudices

But to entrust scientists with dividing up resources for research once they were no longer focused on a common goal seemed to many policy-makers at the time to invite the kind of partisanship that could easily compromise the research process. Specifically, self-serving judgements that scientists routinely suppress when weighing the merits of completed work might re-emerge when evaluating the prospects of work yet to be done. Such scientific outsiders as Faraday, Darwin and Einstein would have probably seemed too radical and unproven to receive peer-reviewed funding, yet their work was readily published by peers upon completion.

As it turns out, concern about élitism

has been well-placed. In recent years, 50 per cent of US federal science funding has gone to 33 out of more than 2,000 institutions of higher education, virtually the same 33 that receive the lion's share of corporate sponsorship². One way of preventing this situation would have been to follow the advice of Frank Jewett, former president of the National Academy of Sciences, and establish a system of checks and balances — one modelled on the US Constitution itself — to prohibit the same scientists from serving on the governing boards of the main journals, professional societies and private and public funding bodies³. In effect, a wider cross-section of scientists would govern their fields, creating a broader sense of scientific 'peerage'. This would, in turn, counteract the so-called principle of cumulative advantage, science's interlocking reward structure that allows the rich to grow richer while the poor get poorer⁴.

More radical proposals would have banished peer review from the funding arena altogether. Peacetime science policy would have resumed the comprehensive welfare programme initiated by the New Deal. Edwin Land, the inventor of the Polaroid camera's self-developing film, called for the federal government to grant scholarships to science students as seed money for potentially lucrative inventions. Others suggested an 'affirmative action' programme for traditionally disadvantaged regions whereby they would receive disproportionately more funds than regions that have traditionally benefited from private endowments. There was even talk of a national service programme that would have set aside scholarships for clever students from impoverished backgrounds to matriculate at the major universities, on the condition that they would then staff technical training and research facilities in economically underdeveloped regions⁵.

All these proposals merit a second look today. Especially noteworthy at the time was the claim that science was suitable for redistributive policies precisely because its modes of enquiry could, in principle, be understood by everyone. Any differences in scientific performance between universities or regions could be attributed to differences in the quality of the training and research facilities at their disposal — not the quality of the people staffing them⁶. This analysis proved persuasive just as long

as the fast-paced, high-tech research that characterized the Second World War was considered an aberration that would be redressed in peacetime by a science policy sensitive to the rate at which scientific innovation can be assimilated by society at large. Unfortunately, a Cold War was declared soon after the Allied victory, and this required the rapid deployment of technically trained personnel, a policy that clearly favoured scientists already at the major research centres.

New battleground

We are on the verge of making the same mistake again, as 'economic competitiveness' has become the ersatz battleground in the post-Cold-War era. The United Kingdom has already taken its place on the front line by subsuming the Office of Science and Technology in the Department of Trade and Industry. Not surprisingly, this has been followed by a spate of calls for the state to concentrate its research funding on the universities best equipped to make the biggest international splash. The most recent and significant call has come from the National Academies Policy Advisory Group (NAPAG), a consortium of the country's most venerable scientific societies. The NAPAG estimates that the number of British universities that are internationally competitive over a broad range of fields is 15 out of around 100. It recommends that research funding should be restricted to the most competitive departments, thereby openly endorsing the principle of cumulative advantage⁷. But will this prove to be a sound strategy?

Ironically, the best argument against making competitiveness the main goal of national science policy comes from Robert Reich, Secretary of Labor in the Clinton administration, which has otherwise been the world's biggest promoter of the competitiveness model. Reich criticizes the model for presupposing that a country's inhabitants necessarily benefit from the economic success of companies with their headquarters nominally within its borders⁸. Applied to science, this means that the number of papers published or patents registered by a country's leading companies and universities does not automatically translate to benefits for either the scientific community or the larger public of that country. That will depend on the full range of its people's skills, attitudes and work

conditions. Arguably, in the past 25 years, Japan and Germany have made better use of the knowledge produced in the United States and the United Kingdom than have the Americans and British. To be sure, scientists already on the 'cutting edge' will always benefit, but that just reinforces the élitism the competitiveness model breeds. A country gains ground on its competitors at the cost of polarizing its own residents.

This problem can be tackled from either the demand or the supply side of science. The demand-driven strategy is to adjust the training, incentives and rewards of scientists (and science students) so they appear more attractive to foreign and domestic companies. Despite the lip service often paid to this strategy, its implementation would be prohibitively expensive. This point contributes to the 'realism' of the NAPAG triage approach to research funding. But even with adequate economic resources and political will, the sheer provision of highly skilled, appropriately motivated scientific professionals does not guarantee their gainful employment in what has increasingly become a buyer's market for brains. As the scientific infrastructure of the developing world grows faster than the rise in labour costs, countries and universities have seen fit to 'outsource' substantial research projects to scientists from these regions. The United States and the United Kingdom have benefited from this 'globalization of R&D' by providing high-grade skills at marginally lower labour costs than Japan and Germany⁹.

False beliefs

But the demand-driven strategy to peacetime science policy suffers from enormous presumptuousness, which in turn reflects the lack of a voice for rank-and-file scientists in setting national science policy. Specifically, the scientific community has never been consulted on its commitment to 'competitiveness' as a goal of its activities, be it defined in military, industrial or even scholarly terms. More generally, whereas the scientific beliefs of nonscientific adults and pre-college science students have been extensively and periodically surveyed, comparable data are lacking for scientists themselves. The NSF's biennial *Science and Engineering Indicators* contains country comparisons of the educational attainments, employment categories, income levels and career patterns of scientists — but nothing about their attitudes toward their work and whether these are adequately represented by their professional spokespersons and policy-makers.

The first step toward a supply-driven response to the competitiveness model of science policy would be just such a periodic inventory of the scientific beliefs of the scientific community, including tertiary-sector science students. Without such systematic information, even leaders of the scientific community have been prone to stereotype

their constituency. A striking example is Leon Lederman, who in 1991, as president of the American Association for the Advancement of Science, issued a "cry of alarm" calling for the federal government to double its non-defence spending on science. To make his case, he restricted himself to surveying the opinions of tenured and untenured faculty members at the 50 leading research universities in the United States¹⁰. Before it was disbanded by the Republican-controlled Congress, the Office of Technology Assessment had begun to take stock of science's missing voices by drawing attention to statistical data on groups that do not fit the stereotypical scientist: graduate students, women, ethnic minorities and the 20 per cent or more of the scientific workforce that is migratory and contract-based, otherwise known as the 'unfaculty'¹¹. Although these data only scratched the surface, they showed that the scientific community has not been uniformly sold on the idea of competitiveness as a peacetime goal.

More indirect support that scientists would welcome a New Deal for science policy comes from a Carnegie Foundation survey on the effects of the research university model on academic work conditions across the entire tertiary sector, including private and public universities and liberal arts and community colleges¹². The imperative to become more 'research competitive' was shown to be a pervasive source of stress in academics' lives, especially in the sciences, as it drew them away from their preferred image of scholarship as educational, open to interdisciplinary influences and less focused on sheer productivity. The survey found that although academic identity was tied more to disciplinary than to specific institutional affiliation, the type of institution in which academics worked shaped their sense of what constituted appropriate scholarship in their discipline. On that basis, accrediting agencies and professional associations should evaluate and promote academic work in ways that avoid polarizing academic institutions into a two-tier funding structure that inevitably ranks research above teaching and relegates public service to the sub-intellectual¹³. Similar findings in the UK context suggest that the minister for science and technology should be relocated to the Department of Education and Labour.

With the end of the Cold War, physics is probably the discipline with the strongest pressure to treat its degree-holders as stakeholders. In January 1991, the Council of the American Physical Society adopted a public position on funding priorities for the first time in its 93-year history, a position that favoured broadly based physics research over the more glamorous but élitist Superconducting Super Collider¹⁴. Three years later, two write-in candidates were elected to the council who, reflecting on the fact that PhD production in physics was

still growing at Cold War rates of 10 per cent a year, called for "limit[ing] the growth of the physics community in the USA to sustainable levels" and "help[ing] young scientists who wish to make a transition out of physics"¹⁵. This has, in turn, sparked a more general debate about whether science degree programmes are broad enough to enable young scientists to work in environments that are increasingly interdisciplinary and nonacademic¹⁶.

Full constituency

Today's problems often rest on maintaining distinctions that previously did not make a difference, and no one was the worse for it. The New Deal approach to national science policy did not presume as strong a distinction between scientists and nonscientists as operates today in, say, the contrast between 'peer reviewed' and 'earmarked' research. Rather, the idea was to extend scientific peerage, not simply by recruiting more scientists, but by ensuring that science's full constituency was represented on governing boards and policy forums. Were this principle in operation today, it is unlikely that we would witness the repeated complaints that federal science panels are 'biased' and 'narrow'¹⁷. Public involvement would not seem so alien to scientific autonomy if a wider range of scientists sat on such panels in the first place. In the twentieth century, élitism in science has been driven primarily by military preparedness and, more recently, economic competitiveness. The twenty-first century offers an opportunity for science to regain the lost legacy of this century, whereby science would be driven by the entire community of enquirers. □

Steve Fuller is in the Department of Sociology and Social Policy, University of Durham, Durham DH1 3JT, UK.

1. Mills, C. *Science* **107**, 127–130 (1948).
2. *Federally Funded Research: Decisions for a Decade* 263–265 (US Office of Technology Assessment, 1991).
3. US Congress House Committee on Interstate and Foreign Commerce *Hearings on the National Science Foundation* 80th Congress, 1st session, 73–76 (March 1947).
4. Merton, R. *The Sociology of Science* 439–459 (Univ. Chicago Press, 1973).
5. Kevles, D. *Isis* **68**, 19 (1977).
6. US Congress Senate Committee on Military Affairs *Hearings on Scientific and Technological Mobilization* 78th Congress, 1st session, 67–72 (April 1943).
7. *Times Higher Edu. Suppl.* **11** (19 April 1996).
8. Reich, R. *The Work of Nations* 122–125 (Random, New York, 1992).
9. *Science and Engineering Indicators* 10th edn, 110 (US National Science Board, 1991).
10. Lederman, L. *Science: The End of the Frontier?* (American Association for the Advancement of Science, 1991).
11. *Federally Funded Science: Decisions for a Decade* 205–230 (US Office of Technology Assessment, 1991).
12. Boyer, E. *Scholarship Reconsidered: Priorities of the Professoriate* (Carnegie Foundation, 1990).
13. Fuller, S. in *Scientific and Technical Communication: Theory, Practice and Policy* (ed. Collier, J.) (Sage, in the press).
14. *Federally Funded Research: Decisions for a Decade* 159, n40 (US Office of Technology Assessment, 1991).
15. Levine, Z. *Biographical Information and Candidate Statements for the 1994 Election of Councilors to the American Physical Society* (American Physical Society, 1994).
16. Holden, C. *Science* **270**, 121–145 (1995).
17. Cordes, C. *Chron. Higher Edu.* **A26** (8 March 1996).

Energetic developments in fracture

Michael Marder

Cracks travel more slowly than standard theory predicts, and finding out why may lead to safer materials and engineering. The answer appears to be in the way a speeding crack branches.

WHAT drives physicists to study cracks? There is certainly some attraction in being able to tell the children one is getting paid to break things. There is also a perverse pleasure in learning the physical laws underlying irreversible change, decay and destruction. Eran Sharon, Steven Gross and Jay Fineberg¹ have just given an answer to an old and deceptively simple question, "How fast do things break, and why?"

The first scientific attempts to answer this question go back to research by Hubert Schardin and Wolfgang Struth in 1937, who used sparks to take photographs of cracks in less than one millionth of a second². They concluded that "the maximum velocity of propagation of glass fractures is to be considered a physical constant", and they measured crack speeds that were approximately a quarter of the speed of sound in many different glasses.

Some time after these experiments came theories of crack motion, which firmly insisted that cracks should move at about twice the speed observed. The classic theory³ left little room for uncertainty. Cracks leave two new surfaces behind them, and the speed at which vibrations travel across a free surface is the Rayleigh wave speed — the speed of sound produced when one raps one's knuckles on the top of a table, or the speed at which earthquakes travel on the surface of the Earth. Cracks, said the theory, should move at this Rayleigh wave speed too — but it does not happen. In Plexiglas, for example, the Rayleigh wave speed is around 1,000 metres per second, but cracks never exceed 600 metres per second. This discrepancy was widely

acknowledged to be rather puzzling, but there was the consolation that it was probably rather unimportant. After all, if a wing is falling off an aeroplane, who has time to wonder whether the crack is moving at 600 or 1,000 metres per second?

In 1991, Fineberg, Gross, Harry Swinney and others developed a method of looking at the motion of cracks, by depositing a very thin layer of aluminium on the surface of a Plexiglas or glass sam-

and accelerate to the proper speed.

The most important clue was found in velocity traces, such as the one on the left of Fig. 1. After accelerating rapidly in its youth, a crack driven hard enough suffers a mid-life crisis at a speed of 340 metres per second, and the smooth motion of the earlier times degenerates into very rough oscillations in velocity.

A large number of different theoretical ideas was proposed to explain these

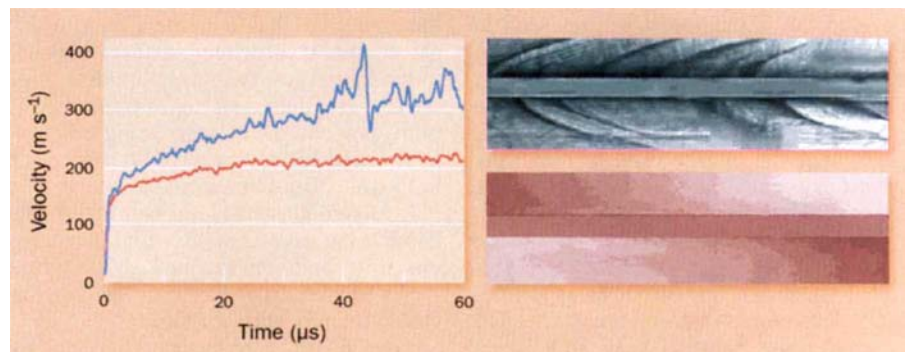


FIG. 1 A crack in Plexiglas, travelling at less than the critical velocity of 340 m s⁻¹, moves calmly and leaves smooth surface in its wake (bottom right). Above this velocity, the crack bucks and plunges, leaving a branching structure beneath the surface (top right). Left, Crack velocity versus time for two experimental runs. (Data courtesy of Sharon, Gross and Fineberg.)

ple, and then monitoring the electrical resistance of the aluminium as a crack raced through it. This technique made it possible to measure the velocity of a crack on timescales much shorter than a millionth of a second, hundreds of thousands of times in succession. With this abundance of new information came many clues about the crack's stubborn refusal to obey a perfectly good theory

experiments. However, one set of ideas is particularly supported by the recent work¹. From this point of view, one regards a crack in its early stages as a thin blade, one atom wide, slicing through a piece of brittle material. The effective sharpness of the blade depends upon the power with which it presses through material: press too hard and it blunts, presenting enormous resistance to speeding

Planes and boats and sticky ends

THE scientific study of breakage originates with Galileo. His *Dialogues on Two New Sciences* begin in a shipyard, where the questioners ask what is "the resistance of solid bodies to separation", and proceed to lay out a sequence of problems to occupy mechanics and solid-state physics for the next three hundred and fifty years. Progress has usually been spurred by disasters. In 1919, a molasses tank 50 feet high and 90 feet wide burst in Boston, killing twelve people and sev-

eral horses. The court auditor concluded that the only rock to which he could safely cling was the obvious fact that at least half of the scientists must be wrong. Better understanding became particularly crucial during the Second World War, when demands for rapid production led to all-welded ships, and no fewer than 88 examples of the first version of the 'Liberty ship' freighter broke beyond repair, sometimes while sitting in port. New methods of testing materials and procedures for designing

ships soon emerged from these failures. Several aeroplane crashes in the following decade were attributed to cracks, which led to systematic inspection procedures. If an engineer is told how much energy it takes to make a crack move forward, he or she can now design ships and planes that will be safe if properly inspected. However, the problem of understanding why a crack needs a certain dose of energy to move, and of calculating when it will do so, is often still mysterious. M. M.

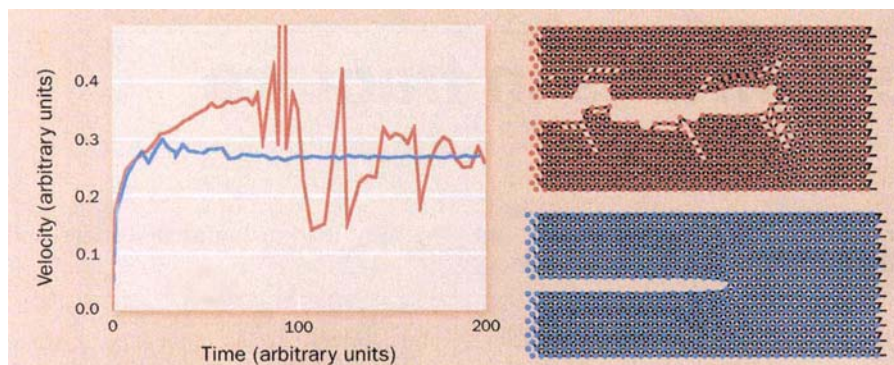


FIG. 2 Simulations in an atomic model of fracture show an instability agreeably similar to experimental observations shown in Fig. 1. Velocities from two crack simulations are plotted on the left, and snapshots of corresponding atomic locations are shown on the right.

up further. Both idealized calculations^{4,5} and more realistic simulations^{6,7} showed that this could occur by the generation of side-branches to the main crack (Fig. 2).

Last autumn, Sharon, Gross and Fineberg⁸ showed that cracks in Plexiglas develop the branching structures seen in Fig. 1. It began to seem plausible that an explanation for the sluggish behaviour of cracks was within reach. However, one part of the argument was still missing: proof that the energy used to make the branches could fully account for the energy robbed from crack acceleration.

This full accounting has now been made. Sharon, Gross and Fineberg carefully looked beneath the surfaces left behind, determined the total energy needed to create the ramified branching struc-

tures they found there, and compared this energy with the total energy that fracture theory claims must be rushing in to the crack tip. The two agree, and the energy needed to create the many subsurface branches is nearly ten times larger than the energy needed to create a smooth single crack. A consistent picture of crack motion emerges. When small amounts of energy are delivered to a crack tip, it plunges forwards, cleanly slicing material in two and leaving perfectly flat surfaces in its wake. But once a critical energy flux is exceeded the crack tip can no longer handle the responsibility, and becomes unstable, building increasingly ramified zones of damaged material beneath the visible fracture surface.

One might wonder whether under-

standing such details of the fracture process could have practical consequences. The prospects are not so entirely discouraging as the aside about the aeroplane wing above would suggest. Indeed, aeroplane safety provides a case in which improved understanding of crack dynamics might be important. Aircraft can tolerate large holes, and still land safely. If a hole does begin to form, the crucial question is whether the cracks responsible will come to rest, or whether they will continue to run until damage is irrevocable. If the current understanding of what sets the speed of cracks can be turned into new ideas on how to make them stop, this research will work its way out of the realm of curiosity and into stronger and safer materials. □

Michael Marder is in the Department of Physics and the Center for Nonlinear Dynamics, University of Texas at Austin, Austin, Texas 78712, USA.

1. Sharon, E., Gross, S. & Fineberg, J. *Phys. Rev. Lett.* **76**, 2117–2120 (1996).
2. Schardin, H., Mücke, L., Struth, W. & Rhein, W. A. *The Glass Industry* **36**, 133–138 (1955).
3. Freund, L. B. *Dynamic Fracture Mechanics* (Cambridge Univ. Press, 1990).
4. Marder, M. & Liu, X. *Phys. Rev. Lett.* **71**, 2417–2420 (1993).
5. Marder, M. & Gross, S. *J. Mech. Phys. Solids* **43**, 1–48 (1995).
6. Xu, X. P. & Needleman, A. *J. Mech. Phys. Solids* **42**, 1397–1434 (1994).
7. Abraham, F. F., Brodbeck, D., Rafee, R. A. & Rudge, W. E. *Phys. Rev. Lett.* **73**, 272–275 (1994).
8. Sharon, E., Gross, S. & Fineberg, J. *Phys. Rev. Lett.* **74**, 5096–5099 (1995).

POPULATION ECOLOGY

When three is a crowd

Michael E. Hochberg

MUCH of the edifice of community ecology is built upon the study of interactions involving pairs or small webs of species. This is perfectly sensible, for communities are in many ways akin to clocks: to understand what makes a clock tick, a good starting point is to fiddle with small ensembles of interacting cogs. Employing such an approach, Begon, Sait and Thompson (page 311 of this issue¹) show how the dynamics of simple species webs are more than just the apparent sum of their parts.

Begon *et al.* conducted their experiments on the Indian meal moth, *Plodia interpunctella* (a stored-products pest) and two of its natural enemies, the parasitoid wasp, *Venturia canescens*, and a granulosis virus. This small community has a number of advantages: the system fits in a small container and has a reasonably short generation length, the artificial diet is easily removed and replaced, and the three populations are straightforward to quantify.

Plodia is already known^{2,3} to exhibit so-called 'generation cycles' when kept free

of its parasitoid and virus. In generation cycles, many or most of the members of a population are in the same age class at any given point in time. This translates into peaks in adult abundance at intervals of approximately one generation. As the generations pass, individual-to-individual differences in development and reproductive rates could blur these peaks. But, because food is limiting in this semi-closed experimental system, and because ensuing competition is most severe for the youngest larval stages, much of the variation in age-structure is truncated from behind in every generation. This truncation tends to cancel the effect of the natural blurring: the cycles are able to maintain their cohesion for at least 15 generations.

Given the moderate virulence of the *Plodia* granulosis virus to the youngest larval stages (lethally infected specimens perish on average after about two weeks), its addition to these otherwise harmonious experimental boxes could be expected to throw the host dynamics completely out of

gear. In fact, as the same authors discovered in previous work², the virus enhances the discreteness of the generation cycles. The minor impact of the virus on host population levels, and the very small prevalences of infected hosts, suggest a limited role for the pathogen in driving the host's dynamics — a limited role which is nevertheless sufficient to refine the intrinsic generation cycles.

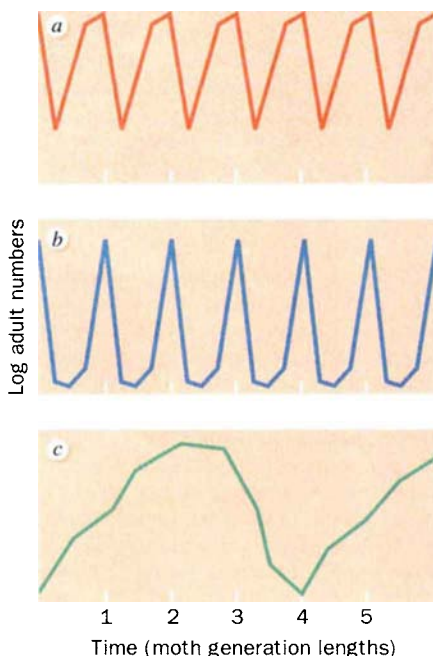
Host generation cycles are also observed, albeit for a different reason, when the parasitoid is added to the *Plodia* boxes³. In contrast to the virus's finesse in honing the intrinsic cycles, the parasitoid appears to induce them itself. In this recent work, Begon and colleagues show that system traits⁴ conducive to such cycles do indeed exist in the *Plodia*–*Venturia* interaction.

The key point is that in both single natural-enemy treatments the generation cycles are fairly regular through time; that is, average moth abundance changes little generation after generation. Does confining both natural enemies to the same box change the dynamics? In their latest study¹, Begon *et al.* find that it does, and dramatically so (see figure). The cycles, although still evident, now occur with a period of three to four host-generation

lengths. Of a total of eight boxes, only four lasted two or more host cycles; four others made it through only one cycle. In all but one of the boxes, the three species died out (the other box being terminated for other reasons). So adding a second natural enemy to the boxes destabilized the system, and led to its inevitable collapse.

The non-equilibrium dynamics of these doomed systems are some cause for excitement. Mathematical models⁵⁻⁷ produce their likes, but they have never been so convincingly demonstrated until now. Although the *Plodia* boxes are a far cry from real communities, they do serve as a promising halfway house between abstract theory and the real world.

To better understand the population ecological vehicles behind these results, first consider each single natural-enemy interaction in isolation. Their persistence is promoted, but not ensured, by unequal risks in mortality imposed by the natural enemy in question. In the case of virus addition alone², different pieces of diet carry different virus loads, meaning that hosts feeding on sparsely contaminated patches will be less prone to infection than those consuming pathogen-rich ones. In the case of parasitoid addition alone³, host vulnerability depends on the position of

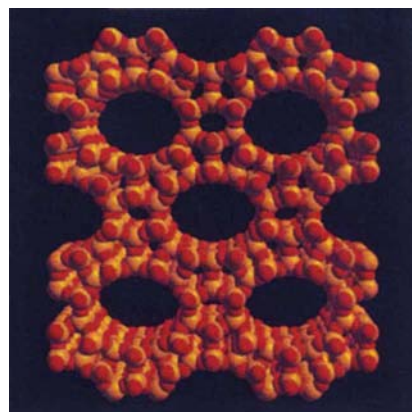


Simplified summary of the results of Begon and colleagues¹⁻³. a, Experimental populations of the Indian meal moth *Plodia interpunctella* exhibit cycles with a one-generation period when confined with a granulosis virus. b, Cycles with the same period, but much more intense troughs, occur when the moth and a parasitoid wasp, *Venturia canescens*, are confined together. c, The picture changes radically when the wasp is added to experimental boxes in which the virus is already present. The cycle is stretched to a period of roughly three to four generations, and the host is eventually driven to extinction, followed by its natural enemies.

Zeolites reach fourteen at last

SYNTHETIC chemists love crystals with holes in. The aluminosilicate zeolites are particularly useful, because their molecular-scale pores absorb organic molecules selectively and the strongly acidic pore walls mean the material acts as a catalyst. Natural zeolite minerals were discovered in the 1800s, and since the 1930s many synthetic versions have been constructed, but all of these contain pores made from rings of no more than 12 aluminium or silicon atoms (each bonded to four oxygen atoms in a tetrahedral arrangement). It is hard to dissuade atoms from minimizing their total energy by packing closely together.

Now this limitation has been overcome, with the material whose crystal structure is shown on the right. On page 295 of this issue, Freyhardt *et al.* describe how they made this zeolite, with its elliptical 14-tetrahedral-atom pore opening, by baking a mixture of



chemicals at 175 °C for two days, bathing the result in hydrochloric acid and then drying it out. The pores in this material are just 0.75 by 1 nm, but they should let larger hydrocarbons through than any other aluminosilicate zeolite, and so allow a range of reactions not possible before. S. B.

the host within the artificial diet — deeply feeding hosts are less likely to be found by roaming parasitoid adult females than surface-feeders. In the two instances, host vulnerability also varies with age, in that early larval stages of the host are more vulnerable to the virus, whereas later ones tend to be sought by the parasitoid.

Now take the full three-species system. The two natural enemies can combine to destabilize the system when the distributions in mortality they each impose differ sufficiently from one another: these can be differences in time or space, or according to host age or phenotype. What is important is that the natural-enemy conglomerate sufficiently imposes itself on the host population (that is, there are few escapees). In the *Plodia* system, larvae escaping viral infection in the first stages of development will always be differentially more prone to attack by the parasitoid in the later stages. So the fundamental difference of the mortality distributions is according to host age.

The intrigue of clever population dynamics aside, Begon and colleagues' study is a poignant illustration of hard upper-limits to natural-enemy diversity: too many natural enemies with overly different exploitation patterns is not a good recipe for maintaining high species richness. These results seemingly run counter to those of Tilman *et al.*⁸, published in February, who showed how the diversity of plants on North American prairies is linked to their productivity and sustainability. But the two conclusions are not really as much at odds as they might seem, in that the nutrient pools adjudicating plant competition in Tilman and colleagues' field plots are evidently more resilient to diverse exploita-

tion strategies than are the moths to the mortalities imposed by their laboratory-bound assailants. Further studies will be necessary to say something more definite about the association between the diversity of natural-enemy complexes and community stability.

The surprise player in the *Plodia* experiments is undoubtedly the granulosis virus. That an organism with such low prevalence (under 5 per cent) can have such far-reaching implications for the structure of a simple community is a potent demonstration of how minor actors in one community can be major ones in another. Arguably, the role of pathogens in community ecology has remained elusive because many of us simply wish to ignore them — they are difficult to observe, to manipulate and to quantify. The elegance with which Begon *et al.* have brought *Plodia* virus to the fore should inspire similar investigations of the effect of pathogens in other systems. □

Michael E. Hochberg is in the Laboratoire d'Ecologie, Université Pierre et Marie Curie, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris 05, France.

1. Begon, M., Sait, S. M. & Thompson, D. J. *Nature* **381**, 311–315 (1996).
2. Sait, S. M., Begon, M. & Thompson, D. J. *J. Anim. Ecol.* **63**, 861–870 (1994).
3. Begon, M., Sait, S. M. & Thompson, D. J. *Proc. R. Soc. B* **260**, 131–137 (1995).
4. Godfray, H. C. J. & Hassell, M. P. *J. Anim. Ecol.* **58**, 153–174 (1989).
5. Hassell, M. P. & May, R. M. *J. Anim. Ecol.* **55**, 923–940 (1986).
6. Hochberg, M. E., Hassell, M. P. & May, R. M. *Am. Nat.* **135**, 74–94 (1990).
7. Wilson, H. B., Hassell, M. P. & Godfray, H. C. J. *Am. Nat.* (in the press).
8. Tilman, D., Wedin, D. & Knops, J. *Nature* **379**, 718–720 (1996).

Stonehenge for the 1990s

Clive Ruggles

LATE last year English Heritage published, in a single volume, full reports on the twentieth-century excavations at and around Stonehenge¹. This has been long awaited by archaeologists. Until the book's appearance, details of the two major campaigns of excavation at the monument — those of Lt-Col William Hawley in the 1920s, and the work of Richard Atkinson, Stuart Piggott and John F. S. Stone in the 1950s and early 1960s — had remained inaccessible. Their availability removes a considerable hindrance to any attempt to understand and interpret the sequence of events at Stonehenge in the wider context of Wessex, an area covering parts of Wiltshire, Dorset, Hampshire and Berkshire, in the Middle-Late Neolithic and Early Bronze Age (roughly 3000–1500 BC).

A meeting* in London had the object of complementing the English Heritage volume and of airing themes not discussed within it. The event took place amongst renewed hopes that a tunnel may now be built to carry the A303 trunk road, which passes within 200 m of the site and is due to be widened, underneath the Stonehenge area. This would return the most celebrated prehistoric monument in the world to rightful prominence within a largely unspoiled landscape, where it could be better enjoyed, appreciated and studied by both public and professionals.

Stonehenge is actually not one monument but a series of them, continually remodelled and reworked over many centuries. One of the most exciting recent developments is the use of Bayesian statistics rather than classical methods to analyse over 50 radiocarbon dates from various features at the monument (A. Bayliss *et al.*, English Heritage). Made practical by the advent of the 'Gibbs sampling' technique², Bayesian methods permit the analysis to take account of archaeological factors — such as the fact that two uncalibrated dates may derive from different archaeological contexts, one of which is known to overlay the other, and hence must postdate it — and provide tighter chronologies as a result³.

We can now place the earliest datable event at Stonehenge, the digging of the circular ditch and bank, within 50 years of 2950 BC, somewhat earlier than was previously thought. (Three post-holes discovered under the present car park represent

unrelated activities in Mesolithic times, some 4,000 years earlier.) The arrival of the bluestones — blocks mostly of dolerite and rhyolite, weighing up to 4 tonnes — is now dated as early as around 2550 BC, several centuries earlier than previously thought. The huge sarsens, the stones that epitomize Stonehenge, arrived only about a century later.

The geological evidence seems overwhelmingly against the idea that the blue-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Every age has the Stonehenge it deserves, or desires — so said the late Jacquetta Hawkes.

stones, whose source is in the Preseli Mountains of southwest Wales, were transported from there by glaciers in the last Ice Age (C. P. Green, Royal Holloway College; J. Scourse, University College of North Wales, Bangor). This has been a much-debated topic^{4,5}. But, for one thing, isolated glacial erratics are not found in ice-marginal areas, only in highly glaciated areas such as the Lake District in north-west England. So we must accept that the bluestones were transported to Stonehenge by people, and face up to the implications in terms of organization, communications and exchange between communities as far apart as Wessex and South Wales (some 200 km). Non-local stones were also used in other Late Neolithic and Bronze Age stone circles, though nothing on this scale.

Why it was important to bring stones from so far away is an open question, as is the issue of how people achieved the almost unimaginable feat of hauling the sarsens, weighing 25 tonnes or more, over 30 km from the Marlborough Downs in the north. New experiments (J. Richards, English Heritage; M. Whitby, Whitby & Bird, London) show that large sarsen

stones could be pulled up 1-in-20 hills using just 130 people, moving them by as much as 200 m or more in a day. The biggest problem is to overcome static friction in order to get started. The experiments have also demonstrated how a sarsen could have been neatly dropped over into a prepared hole by using a pivoting stone underneath and a small sliding stone or stones on top.

The new archaeological evidence suggests a new chronology, superseding Atkinson's Stonehenge I, II and III (ref. 6). The first Stonehenge (Stonehenge 1) was built in an already thriving Middle Neolithic landscape. Environmental evidence, including analyses of bone, pollen and molluscs, reveals a landscape that gradually changed over the next 1,500 years from a patchwork of woodland and cleared scrubland, with some cultivation, through to almost total use of the land for pastoral and arable farming. This transition was accompanied by considerable social change, with an ever-larger labour force that could evidently be mobilized for large communal projects. A tradition of circular monument building was perpetuated, at first in timber (Stonehenge 2) and then in stone (Stonehenge 3), each stage in the development involving an order-of-magnitude increase in the workload involved. Remarkably, the stone monument was then used, with various structural modifications, for an entire millennium.

What was its purpose? It was clearly a centre for meetings and rituals — for intense ceremonialism. There are vital clues as to the nature and meaning of the events that took place there. Layers of symbolism were evidently encoded in the architecture of the monument, in its size, shape and texture, in the materials used (including a variety of non-local stones), in the positions where commemorative offerings and other deposits were placed, in what happened inside the circles and what was excluded, and in what was seen and experienced by people approaching the monument (especially in the later stages, along the 2.5-km-long avenue). A built-in asymmetry in the sarsen trilithons, constructions of two standing stones and a lintel, may have had particular significance in relation to bodily orientation (A. Whittle, University of Wales, Cardiff). But many aspects of the meaning of Stonehenge still remain elusive.

New fieldwork, together with computer analyses using Geographical Information Systems (D. Batchelor, English Heritage), has permitted analyses of the areas visible from Stonehenge and other monuments in the vicinity, and from which the monuments are visible. They reveal that Stonehenge sits at the centre of a number

* Science and Stonehenge, London, 20–21 March, 1996.

of 'nested bowls', and that visibility considerations influenced the siting of nearby monuments; the Early Bronze Age barrows, for example, were placed in lines along the ridges forming the near horizon so that each could be clearly seen from Stonehenge.

Finally, whatever happened to the astronomy? Jacquetta Hawkes, who sadly died the day before the conference began, once famously said that every age has the Stonehenge it deserves, or desires. The idea that Stonehenge was some sort of astronomical observatory or computer, so popular in the 1960s, is now seen as an artefact of its times — one of the most notorious examples known to archaeologists of an age recreating the past in its own image. Yet the axial alignment upon midsummer sunrise and midwinter sunset does seem to have formed a vital part of the symbolism of the stone phases of the monument; it is strengthened by the discovery in 1979 of a probable companion to the Heelstone⁷, the two stones forming part of a 'corridor' down which the Sun would have shone into the interior of the sarsen circle on mornings around midsummer⁸.

The evidence for other astronomical symbolism, including a possible lunar significance in the earlier phases^{9,10}, is more equivocal. But the idea that the monument may have symbolized cyclical time through alignments on the Sun or Moon, or that astronomical considerations formed part of the sacred principles that helped to structure the Neolithic and Bronze Age landscape around Stonehenge, no longer seems far-fetched. □

Clive Ruggles is in the School of Archaeological Studies, University of Leicester, Leicester LE1 7RH, UK.

1. Cleal, R. M. J., Walker, K. E. & Montague, R. *Stonehenge in its Landscape: Twentieth-Century Excavations* (English Heritage, London, 1995).
2. Gelfand, A. E. & Smith, A. F. M. *J. Am. statist. Ass.* **85**, 398–409 (1990).
3. Buck, C. E. et al. *Antiquity* **65**, 808–821 (1992).
4. Kellaway, G. A. *Nature* **233**, 30–35 (1971).
5. Thorpe, R. S. et al. *Proc. prehist. Soc.* **57**, 103–157 (1991).
6. Atkinson, R. J. C. *Stonehenge*, 3rd Edn (Penguin, Harmondsworth, UK, 1979).
7. Pitts, M. W. *Proc. prehist. Soc.* **48**, 75–132 (1982).
8. Burl, H. A. W. *Wilts. Arch. nat. Hist. Mag.* **87**, 85–95 (1994).
9. Burl, H. A. W. *The Stonehenge People* (Dent, London, 1987).
10. Castleden, R. *The Making of Stonehenge* (Routledge, London, 1993).

SOLAR SYSTEM

The enigma of Jupiter's ring

Gregor Morfill

SATURN has been known for more than three hundred years as 'the ringed planet' (largely through the work of Christiaan Huygens¹), whereas the existence of rings around the other giant gas planets — Jupiter, Uranus and Neptune — has been confirmed only comparatively recently. The reason is visibility: Saturn's ring appears in modern high-resolution telescopes as a huge, majestic object, whereas Jupiter's ring is much less imposing. In fact, it needed nearby observations from the cameras on board the Voyager spacecraft about 16 years ago to confirm its existence^{2,3}. It is about twenty Earth diameters across, and its radial width is about half an Earth diameter — proportions similar to those of a bicycle tyre. In contrast to Saturn's rings, which are made of metre-sized 'boulders', the visible ring around Jupiter is a 'dust ring' — it consists largely of micrometre-sized particles, with diameters comparable to the thickness of a human hair. It is so tenuous that if all the dust were swept up and compacted into a single cubic block of material, it would measure a mere 40 m along each edge.

This tenuous jovian ring has been somewhat of an enigma. After 16 years of research we still do not understand where it comes from or how it works. On page 293 of this issue⁴, Horányi and Cravens propose a solution which finally promises

the long-awaited breakthrough. To appreciate the achievement and to understand the puzzle, let us go back to Voyager's Jupiter fly-by and the subsequent developments in planetary ring science triggered by the discovery.

Immediately after the first observations



The tenuous dust ring of Jupiter, photographed by the Voyager 2 spacecraft as it flew by in 1979.

of this tenuous ring it became clear that it could not be a stable, long-lived structure^{5–7}. The orbits of small particles are easily perturbed, even by radiation pressure and electromagnetic forces, and the particles then get swept up by the planet and are lost. So there had to be a source.

It was proposed that Jupiter has a ring of larger, kilometre-sized 'moonlets', which are constantly bombarded by small sub-micrometre projectiles. The impacts lead to an 'excavation' or 'mining' of these bodies, the ejecta forming the supply for the tenuous ring. While these new concepts were being formulated, two moonlets, later named Adrastea (radius about 10 km) and Metis (radius about 20 km) were discovered through painstaking analysis of the Voyager images^{8,9}. They are both located near the outer edge of the ring. Assuming, then, that these two moonlets are the tip of the iceberg — the largest of a whole group located at about 1.8 Jupiter radii — we have the source of material required to populate the visible ring.

But where do the projectiles come from? Two possibilities were discussed, particles from the interplanetary cloud that is responsible for the zodiacal light⁶ (a dim glow best seen about one hour after sunset or before sunrise), or small dust particles ejected into Jupiter's magnetosphere by volcanic activity on the moon Io⁵. It would seem easy, at first sight, to check the plausibility of the interplanetary source. The mass distribution and fluxes are known¹⁰, the physics of high velocity impacts has been studied in the laboratory¹¹, and the optical properties of particles can be calculated and compared with the Voyager observations of the ring.

There the problems start. Knowing the projectile particle fluxes doesn't help if we don't know the target area presented by the moonlets. Worse, to calculate the spatial structure of the ring and then compare it with the observations, we need to know the size distribution of the ejecta and the relevant particle transport and destruction mechanisms. Whereas radiation pressure effects are reasonably well known, the strengths of plasma drag and of electromagnetic forces are not. The same is true for losses through sputtering (erosion by ion bombardment), collisions with other dust particles, and atmospheric absorption.

While more and more insight was gained into the complexity of particle charging and transport in planetary magnetospheres, progress in the analysis of Voyager data revealed that the ring is structured: it consists of a main ring, as described above, with a scale height of about 300 km, which at its inner edge expands into an even fainter torus¹² with a thickness of about 10,000 km. At this stage most modellers threw in the towel.

So what is the exciting new aspect of Horányi and Cravens' work? They started off with the basic impact model developed in the 1980s, and improved the dynamical modelling of particle trajectories. Proposing that the plasma environment in the

jovian ring regime is dominated by ionospheric photoelectrons and ions, they arrive at the astonishing result that the ejecta particles have lifetimes of less than a year — more precisely, as short as $t \approx 100 a^3$ days, where a is the particle's radius in micrometres. Older estimates based on sputter erosion had yielded lifetimes of 100 years or more.

The authors were then able to deduce the size distribution of the ejecta particles and find that their production rate is proportional to a^γ , with $\gamma \approx -5.5$ providing a good fit to both the optical depth of the main ring and the thickness of the faint torus. In addition, the density distribution in the ring is in agreement with the light scattering observations¹². All known features of the jovian ring are explained.

Horányi and Cravens do not discuss one logical consequence of their analysis, however. Each hypervelocity impact produces a spray of ejecta with a characteristic size distribution that depends on the mass and the velocity of the impacting particle¹¹. If the mass spectrum of the impacting particles is a power law, then the source spectrum of ejecta particles feeding the ring must be the same power law. This implies that the size spectrum of the projectiles is very steep, proportional to $a^{-5.5}$. From the observations and the physics of high-velocity impacts¹¹, the size range of projectiles is $0.01 \leq a \leq 10 \mu\text{m}$. But in this size range, interplanetary grain size distributions¹⁰ have a power law proportional to $a^{-2.5}$. By no stretch of the imagination is it possible to make this compatible with Horányi and Cravens' careful and highly sophisticated analysis.

This appears to leave no alternative (barring as-yet-unknown processes) to the early suggestion⁵ that volcanic ash ejected by the moon Io is the source of bombarding particles responsible for the tenuous jovian ring. If this is so, then the steep size spectrum must be related to that of the particle flux from the volcanic plumes, which is known¹³ to peak at about $0.01 \mu\text{m}$. A fascinating deduction. □

Gregor Morfill is in the Max-Planck-Institut für extraterrestrische Physik, D-85740 Garching, Postfach 1603, Germany.

1. Huygens, C. in *Oeuvres Complètes de Christiaan Huygens* Vol. XV, 209–353 (The Hague, 1925).
2. Smith, B. *et al. Science* **204**, 951–971 (1979).
3. Owen, T. *et al. Nature* **281**, 442–446 (1979).
4. Horányi, M. & Cravens, T. E. *Nature* **381**, 293–295 (1996).
5. Morfill, G., Grün, E. & Johnson, T. V. *Planet. Space Sci.* **28**, 1101–1117 (1980).
6. Burns, J. *et al. Icarus* **44**, 339–360 (1980).
7. Consolmagno, G. J. *Nature* **285**, 557–558 (1980).
8. Jewitt, D. C., Danielson, G. E. & Synnott, S. P. *Science* **206**, 951 (1979).
9. Synnott, S. P. *Science* **210**, 786–788 (1980).
10. Leinert, C. & Grün, E. in *Physics of the Inner Heliosphere* (eds Schwenn, R. & Marsch, E.) 207–275 (Springer, Berlin, 1990).
11. Grün, E. *et al. Icarus* **44**, 326–338 (1980).
12. Showalter, M. R., Burns, J. A., Cuzzi, J. N. & Pollack, J. B. *Icarus* **69**, 458–498 (1987).
13. Collins, S. A. *J. geophys. Res.* **86**, 8621–8626 (1981).

Inside an intron invasion

Anna Marie Pyle

IN addition to the gradual changes brought about through mutation and selection, there are other processes that mix up the blueprint for an organism, like the apparent wandering or insertion of long DNA sequences, known collectively as mobile genetic elements¹. Because certain types of mobile elements encode RNA molecules with innate catalytic activity, it has long been debated whether catalysis by these RNA products has anything to do with the apparent mobility of their genes. On page 332 of this issue², Yang and co-workers now show that group II intron RNA can insert itself directly into double-stranded DNA, and in doing so they have stumbled upon an unexpected pathway for intron mobility and identified an important connection be-

tween DNA mobility and RNA catalysis.

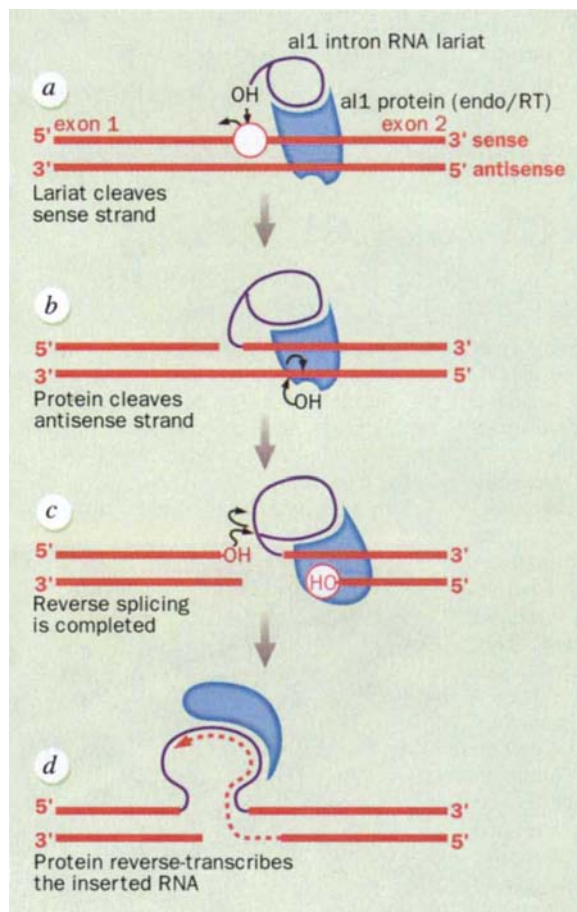
Group II introns are one of four types of intervening sequences, or introns, located between the protein-coding sequences, or exons, of eukaryotic genes. By folding themselves into catalytically active structures, group II introns transcribed from DNA into RNA excise themselves from precursor messenger RNA through two transesterification reactions in a process known as self-splicing³. Although group II intron RNAs can catalyse these and other reactions in the absence of proteins *in vitro*, during the catalysis of attack and insertion into DNA, the introns get some help.

The aI1 and aI2 introns from yeast mitochondria each encode a protein that serves as an essential collaborator in the

catalysis of DNA cleavage and invasion (reverse-splicing into DNA): each protein associates specifically with its own intron, forming a ribonucleoprotein particle. Although it is common to assume that one component of a ribonucleoprotein particle is the catalyst and the other is a structural constituent, in this case it appears that functional active sites are important on both the RNA and protein for reverse-splicing into DNA.

Previous models put forward to explain intron mobility did not necessarily require active participation by transcribed intronic RNA¹. Reverse transcription from unprocessed RNA into DNA could conceivably generate a double-stranded copy of the intron that might be stably incorporated during recombination. Alternatively, as group I and group II introns catalyse reverse splicing into other RNA molecules^{4,5}, it has been proposed that intron migration could occur when intronic RNA inserts itself into foreign RNAs, followed by reverse transcription to a DNA copy and integration into the genome⁶.

That mobility might involve catalysis by the RNA itself, and intron insertion directly into DNA has re-



Insertion of a group II intron into double-stranded DNA, as demonstrated by Yang *et al.*². Lariat intron RNA is a reaction product of group II intron self-splicing which can attack the sense strand of target DNA (a). The aI1 protein contains both endonuclease (endo) and reverse transcriptase (RT) domains, and it attacks the complementary antisense strand of the DNA (b). After the second stage of insertion (c), the intron becomes covalently attached to both halves of the sense DNA strand (d). The aI1 protein then makes a DNA copy of the inserted intron by priming with antisense DNA. Thick red lines, the two DNA strands at the insertion site; thin purple line, aI1 intron RNA; dashed red line, DNA copy of the intron.

cently been indicated by investigations of intron 'homing' into intronless variants of a gene^{7,8}. Zimmerly and co-workers discovered that the aI2 ribonucleoprotein catalyses double-stranded DNA cleavage at a target site bearing sequences recognised by complementary sequences within the intron RNA. Following DNA cleavage, the aI2 protein reverse-transcribes the intron RNA, using the 3' end of DNA at the cleavage site as a primer⁷. In a subsequent study of this reaction, the authors reported that aI2 RNA actually catalyses cleavage of the DNA sense strand, while the aI2 protein cleaves the antisense strand⁸. The protein also serves to enhance the ribozyme activity of aI2 RNA, enabling it to react at low ionic strength and to prefer DNA substrates.

In their latest work using the aI1 ribonucleoprotein particle², the authors have shown that the intron RNA not only attacks the DNA sense strand and serves as a template for reverse transcription, but that it completely incorporates itself into the DNA through a two-step reverse-splicing reaction (see figure). Thus, a catalytic group II intron RNA can invade a DNA molecule, providing a direct means for intron incorporation into a genome. At this stage, it is not clear how reverse transcripts of this chimaeric product become stably incorporated into a DNA genome, although certain mechanisms of DNA repair and recombination may complete the process.

Although the reactions reported by Yang *et al.* are remarkable, they are consistent with established features of group II intron reactivity. After all, it is known that reverse splicing of group II introns into RNA is readily accomplished^{5,9}, and that group II intron ribozymes can cleave DNA molecules¹⁰. Certain group II intron ribozymes catalyse the cleavage of RNA and DNA with comparable efficiency¹¹. This mechanistic property, together with an apparent lack of contact with particular ribose 2'-hydroxyl groups¹¹, is to be expected for a ribozyme active site that may have evolved to catalyse mobility into DNA.

Similarities in their reaction mechanisms have led to the proposal that group

II introns are evolutionary predecessors of the spliceosome³, a complex assembly of ribonucleoprotein particles responsible for removing introns from precursor messenger RNA in the nucleus. DNA integration by group II introns provides a potential model for intron dispersal — an event that may have been followed by deconstruction of group II introns into modern nuclear introns and spliceosomal RNAs.

Clues for identifying other potential relatives of group II introns may be found in the sequences of their encoded proteins⁷. The aI1 and aI2 protein components each contain reverse-transcriptase domains similar in sequence and function to those found in non-LTR retrotransposons — a separate class of mobile genetic elements¹². The DNA-primed reverse-transcription reaction of the aI1 and aI2 proteins is also reminis-

cent of telomerase, an enzyme that contains an RNA template to help it replicate the ends of chromosomes. This parallel is particularly striking, given that *Drosophila* uses a non-LTR retrotransposon instead of a telomerase to keep its chromosomal termini in good shape.

Distant relatives aside, the primary significance of this work is that it extends the relevance and catalytic repertoire of modern RNA catalysts. It demonstrates that RNA and protein active-sites can catalyse reactions in concert, and that, by applying the individual strengths of each, ribonucleoproteins may promote complex chemical transformations. □

Anna Marie Pyle is in the Department of Biochemistry and Molecular Biophysics, College of Physicians & Surgeons of Columbia University, 630 West 168th Street, New York, New York 10032, USA.

ENERGY METABOLISM

Creatine kinase shapes up

George L. Kenyon

WHEN you need a burst of energy, a creatine kinase (CK) will undoubtedly be involved. These are a family of distinct but related enzymes found in abundance in rapidly respiring tissues of vertebrates, including smooth and striated muscle, heart muscle, brain, sperm and mitochondria. In a landmark paper that appears on page 341 of this issue¹, Fritz-Wolf *et al.* report the very first CK crystal structure. It is of the octameric Mi_b (for mitochondrial)-CK from chicken cardiac tissue, and reveals for the first time some of the key structural features of the catalytic (active) site. This is a notable achievement because the structural solution of a CK had eluded several other crystallographic groups for many years.

The apparent function of CK is to buffer the level of adenosine 5'-triphosphate (ATP, a primary energy source in cells) by the interconversion of phosphocreatine and adenosine 5'-diphosphate (ADP) with creatine and ATP. As such, CK plays a central part in the specific cellular utilization of ATP, and may also be involved in tumour formation² and some signalling pathways³. Levels of CK are commonly monitored clinically in serum following myocardial infarctions and are also used to detect certain forms of muscular disorders.

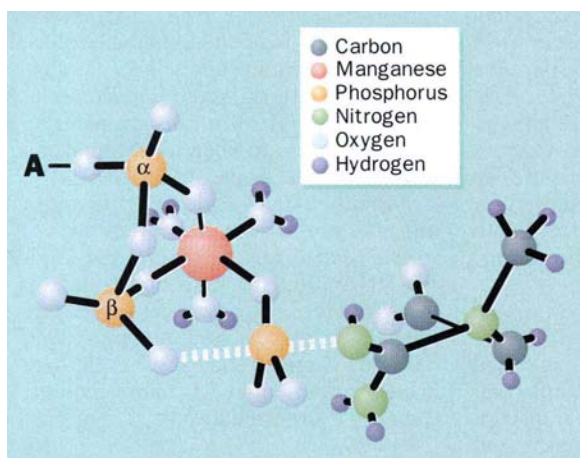
Most of the mechanistic and biophysical studies have been performed on the cytosolic M(for muscle)-CK from rabbit skeletal muscle (RMCK), which is 65 per cent identical in sequence to the chicken Mi_b-CK. Because the key, conserved amino-acid residues have different numbers, the standard notation in this report

will be to designate the Mi_b-CK residues numerically, immediately followed by the RMCK designations in parentheses — for example Cys 278 (282, or 283 to the authors¹, who count from the starting methionine).

The crystal structure of the Mi_b-CK is fully consistent with data from other CKs such as RMCK. Although Fritz-Wolf *et al.* found two somewhat different NaATP binding sites in two distinct crystal forms and no creatine (or analogue) was present in their structures, the nucleotide binding site definitely appears to be located in the central region of the monomer (see Fig. 1a–c, page 342) encompassing, among others, portions of strands β₆, β₁ and helix α₉. Remarkably, the CK structure appears to be unique among kinases, although, as the authors point out, the L₂₈₂GTGLRAGV₂₉₀ sequence motif is commonly found in protein kinases. In fact, GXGXXG, present in this region of RMCK, has been proposed as a putative signature sequence for nucleotide binding sites in general⁴. The ATP is bound with the adenine base in the *anti* orientation relative to the ribose, as had been predicted from results of studies on the nuclear Overhauser effect on MgATP bound to RMCK⁵.

A useful way to examine a new structure is to see how key amino-acid residues that have been identified earlier can interface with the active site. A number of such localizations of residues are noted by Fritz-Wolf *et al.*, namely, Cys 278 (282), Val 232 (236)–Cys 237 (241), Val 272 (279)–Arg 287 (291), Asp 335 (340), Trp 223 (228) and Ala 323 (327). In

1. Lambowitz, A. M. & Belfort, M. A. *Rev. Biochem.* **62**, 587–622 (1993).
2. Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. *Nature* **381**, 332–335 (1996).
3. Michel, F. & Ferat, J.-L. A. *Rev. Biochem.* **64**, 435–461 (1995).
4. Woodson, S. A. & Cech, T. R. *Cell* **57**, 335–345 (1989).
5. Augustin, S., Müller, M. W. & Schweyen, R. J. *Nature* **343**, 383–386 (1990).
6. Müller, M. W., Alimaier, M., Eskes, R. & Schweyen, R. J. *Nature* **366**, 174–176 (1993).
7. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. *Cell* **82**, 545–554 (1995).
8. Zimmerly, S. *et al.* *Cell* **83**, 529–538 (1995).
9. Chin, K. & Pyle, A. M. *RNA* **1**, 391–406 (1995).
10. Möri, M., Niemer, I. & Schmeizer, C. *Cell* **70**, 803–810 (1992).
11. Griffin, E. A., Qin, Z.-F., Michels, W. A. & Pyle, A. M. *Chem. Biol.* **2**, 761–770 (1995).
12. Luan, D. D., Korman, M. H., Jakubchak, J. L. & Eickbush, T. H. *Cell* **72**, 595–605 (1993).



The creatine kinase reaction — putative transition state in the phosphoryl transfer reaction between MnADP (left) and phosphocreatine (right) to form MnATP and creatine. The structure of this activated complex was deduced using electron paramagnetic resonance studies of analogue complexes in the presence of oxygen-17 labelled nucleotides at the α - and β -phosphorus positions¹⁴. A is the adenosine moiety.

particular, the location of Cys 278 (282) merits comment, as its role in CK's mechanism has been rather controversial^{6,7}. It is apparently located at a β -turn⁸ between helix α_9 and strand β_6 (Fig. 1c of the paper) and, as such, looks to be strategically positioned to participate directly in catalysis. Alternatively, when modified by either relatively bulky or charged sulphydryl blocking agents, it can change the active site's geometry or steric access (or both), effectively obliterating catalysis. A more definitive insight into its involvement will have to await solution of a structure in the presence of either creatine or an analogue. In any case, a mutant RMCK with serine rather than cysteine at position 282 has been shown to be at least partially active (relative rate about 1/500 that of wild-type)⁶. Interestingly, Lys 195 in RMCK can be crosslinked with Cys 282 (ref. 9), indicating that they are close together in the structure. This result again is consistent with the M_i -CK crystal structure.

As Fritz-Wolf *et al.* discuss, the function of one of the conserved histidine residues in CK's catalytic mechanism as a possible acid/base catalyst has been the subject of intense scrutiny. Of all of the histidine residues found in the 43 guanidino kinase sequences published to date^{10,11}, only five are conserved — His 92 (96), 102 (105), 186 (190), 229 (233) and 291 (295). All of them have been individually changed by mutation to their asparagine (N) counterparts in RMCK¹¹, each mutant then being over-expressed in *Escherichia coli* and characterized. H105N, H190N and H233N all displayed specific activities similar to those of the wild-type enzyme. Clearly, then, these histidines cannot be actively involved in catalysis. Expressed in a crude extract, H96N also had high specific activity, but the mutant seems to be quite unstable and

has not yet been further purified. Its corresponding M_i -CK residue (His 92) is clearly observable in Fig. 1b of the paper as being near the γ -phosphate group of the bound ATP. The most interesting mutant is H295N which showed much reduced, but still measurable, catalytic activity. His 291 (295) is near the nucleotide-binding pocket in the crystal structure. Its precise involvement in catalysis, however, will have to await more detailed structural information.

These mutational results with conserved histidine residues reinforce lessons that were predicted by Eigen¹² and dramatically emphasized by mutagenesis studies on thymidylate synthase¹³. It is advantageous for enzymes to have evolved to be robust — that is, mutations of residues even at or near the active site do not necessarily obliterate catalysis, but often have either no or only a minor effect. This is probably especially true of enzymes such as kinases, which rely primarily on proximity to achieve their catalytic effects. The main function of the enzyme is evidently to bring the substrates together intimately so that direct phosphoryl transfer can occur between ATP and the chosen substrate. Of course, the kinase provides the appropriate optimal electrostatic environment to promote charge-stabilization in the transition state (see figure). As Fritz-Wolf *et al.* point out, there is an abundance of positively charged residues at or near the nucleotide binding site to help create just such an environment. □

George L. Kenyon is in the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, USA.

1. Fritz-Wolf, K., Schnyder, T., Wallimann, T. & Kabsch, W. *Nature* **381**, 341–345 (1996).
2. Teicher, B. A. *et al.* *Cancer Chemother. Pharmacol.* **35**, 411–416 (1995).
3. Chida, K. *et al.* *Biochem. biophys. Res. Commun.* **173**, 346–350 (1990).
4. Rossman, M. G., Moras, D. & Olsen, K. *Nature* **250**, 194–199 (1978).
5. Rosevear, P. R. *et al.* *Biochemistry* **26**, 5338–5344 (1987).
6. Chen, L. H. *et al.* *FASEB J.* **4**, A2119 (1990).
7. Furter, R., Furter-Graves, E. M. & Wallimann, T. *Biochemistry* **32**, 7022–7029 (1993).
8. Maggio, E. T. *et al.* *J. biol. Chem.* **252**, 1202–1207 (1977).
9. Mahowald, T. A. *Fed. Proc.* **28**, 601 (1969).
10. Mühlebach, S. *et al.* *Molec. Cell Biochem.* **133/134**, 245–262 (1994).
11. Chen, L. H., Borders, C. L. Jr, Vásquez, J. R. & Kenyon, G. L. *Biochemistry* (in the press).
12. Eigen, M. *Chem. Ser.* **268**, 13–26 (1986).
13. Carreras, C. & Santil, D. V. A. *Rev. Biochem.* **64**, 721–762 (1995).
14. Leyh, T. S. *et al.* *Biochemistry* **24**, 308–316 (1985).

Shades of health

FACED with almost any complaint, a doctor's prime wish is that he could have seen it sooner. Early diagnosis is the key to cost-effective medicine. Daedalus now has a diagnostic breakthrough.

A tattoo, he remarks, is a pigment-implant about a millimetre under the skin, among the superficial cells of the dermal capillaries. It makes good contact with the circulating blood. Clinical laboratories these days have hundreds of elementary blood tests, each of which indicates the level of some key metabolite by a simple colour change. Daedalus plans to incorporate them all into a novel diagnostic tattoo.

A litmus tattoo, for example, would reveal blood pH. It would go red if the blood went acid, and blue if it went alkaline. Blood oxygen and carbon dioxide, calcium, potassium and iron, glucose and lactic acid, crucial vitamins and hormones, each could similarly be revealed by a suitable indicator. A thermochromic tattoo could show the body temperature. Piezochromic microballoons, collapsing under pressure to eclipse their surface colour, could reveal blood pressure; they might even pulsate visibly to display the heart-rate.

The diagnostic tattoo will probably be most useful on the inside of the wrist. It will be a grid of little coloured squares, each showing the level of a specific blood component. Levels which vary widely will need several squares, changing colour in sequence as the level rises. The wearer need not bother with the details; he can just check the tattoo as a whole against a standard shade card carrying the 'normal' colour pattern. Any deviation will send him to the doctor at once.

If the doctor then prescribes a drug, he need not worry about the dosage. The patient will simply take another tablet whenever the relevant square of the tattoo drifts towards the 'level too low' shade. In the same way, an alcoholic motorist could 'titrate' himself almost up to the legal blood-alcohol limit, and stop drinking when the warning square on the tattoo begins to change colour.

Sexually active women will use the tattoo to check their changing hormone levels. The monthly player of sexual 'Vatican roulette' will at last know her safe period exactly. The Pill-user will have daily reassurance of its adequacy, while the would-be mother will be instantly advised of her pregnancy.

Sadly, the first and most eager customers for the new tattoo will be hypochondriacs. They will spend hours scanning their tattoos for deviations from the shade card, and will reply to the innocent query 'How are you?' in devastating detail.

David Jones

Visual illusion from running

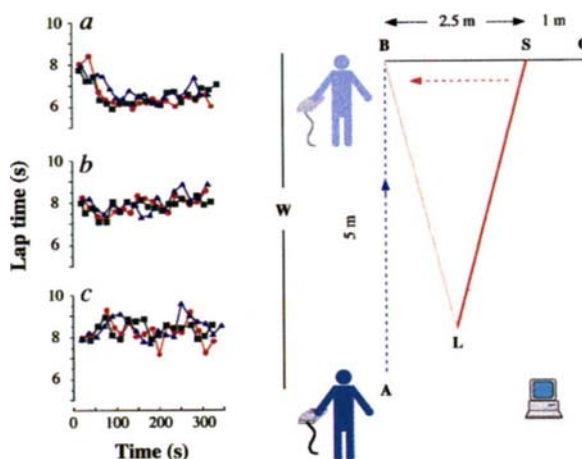
SIR — We have observed a striking visual illusion. When walking on solid ground after more than 10 minutes of fast jogging or running on a treadmill exercise machine, people experience a sensation, lasting 2–3 minutes, of apparently moving at a markedly accelerated pace, as when walking on the passenger-transporting devices (travelators) often found in airports.

We and 14 other subjects (total age range 23–73 years), seven of whom were naive to the expected effect, reported the sensation of accelerated self-motion in comparing their experiences when walking in stationary surroundings before and immediately after treadmill running. In separate experiments, subjects were instructed to maintain a constant 'visual speed' while repeatedly walking a 5-metre course, and were seen to accelerate their pace as the illusion gradually vanished with successive laps (see figure).

We made the following additional observations. First, the effect is strongest after treadmill running with eyes open; we also observed a weak effect after treadmill running while blindfolded, running outdoors, or exercising on a stationary bicycle. Second, occluding peripheral visual information during treadmill running by looking through a restrictive visor reduces the strength of the effect. Third, we observed the illusion when moving forwards, backwards or sideways only after movement on the treadmill in the corresponding manner. Fourth, running with the treadmill close to a textured wall on one side leads to a stronger effect when walking next to a wall on the same side compared with the other side. Fifth, long-term experience of exercise on treadmills reduces the strength of the effect. Sixth, there is no reduction in strength after running with head and elbows stabilized on a shelf mounted above the treadmill. Seventh, active movement is necessary to observe the

effect — subjects report no effect when pushed on a wheelchair instead of walking. And finally, the effect is most compelling when walking on the treadmill itself after it has stopped.

We think the first four observations



Left, quantifying the treadmill visual motion illusion. Successive lap times are plotted as a function of elapsed time for three male subjects (aged 24, 28 and 34 years) who were instructed to maintain a constant 'visual speed' while repeatedly walking a 5-m course. A gradually accelerating walking pace is evident as the illusion decays after running (20 min at 10 km per h) on a treadmill (a), but not after running outdoors (b), or after a control period without prior treadmill activity (c). Right, the test setup. The first lap in each trial was designed to ensure that all subjects started walking at the same pace. A red laser point-source emanating from L was swept at a speed of 8 s per 2.5 m across the wall facing the subject, from O to B. When the spot of light was at S, the subject began walking from A to B, alongside wall W, adjusting his pace so as to meet the spot at B. After the first lap, the spot disappeared at S, at which point the subject would begin each lap. The subject timed the laps by pressing a button at start and finish on the hand-held mouse.

show that the illusion is primarily generated by vision, and results from activity-dependent discrepancies between the observed and expected peripheral optic flow of locomotion. From the fifth observation we conclude that repeated exposure to the adapting conditions reduces these discrepancies. In the sixth, the up-and-down visual motion that normally accompanies running was reduced or eliminated, together with corresponding vestibular signals; we conclude that these are not important for the adaptation. The seventh observation suggests at first that signals initiating active motion are necessary, but their role could be taken by the proprioceptive and efference copy signals actually accompanying locomotion. From the final test we conclude that re-exposure to more of the adventitious sensory signals accompanying the main adapting stimulus heightens the effect.

Two related after-effects have recently

been reported. In one study¹, blindfolded subjects who had jogged for 60 seconds on a treadmill inadvertently advanced when asked to jog in place, still blindfolded, on solid ground. In the other study², subjects walked for 8 minutes on a treadmill which itself was pulled along by a tractor at different speeds; they were then shown a target and asked to walk to it blindfolded on solid ground. When the tractor speed was lower than the treadmill-walking speed, the subjects overshot the target, and when it was higher they undershot it. In these experiments, subjects made biased movement in atypical self-motion situations without any visual input during the test period. Our illusion shows that disturbing the normal relation between self-induced motion and the expected sensory input leads subjects to experience altered motion and misjudge its extent even in normal locomotion with eyes always open.

We believe that all these phenomena are related to those Helmholtz explained as unconscious inference³, and involve recalibration of the mechanisms estimating associations between sensory messages. Changes in these associations are often of vital importance, for they signal changes in the causal net within which we live. One way to detect them sensitively would be to adjust the representational variables continuously, so that they are decorrelated⁴; in the illusion described here, this achieves the same end for variables representing self-motion inputs from two or more sensory modalities as automatic gain control does for a single sensory variable (for example, in light adaptation). Adaptation after-effects, including those discussed here, would then be the outcome of adjustments made during a period of novel sensory experience.

A. Pelah*

H. B. Barlow

Physiological Laboratory,
University of Cambridge,
Downing Street,
Cambridge CB2 3EG, UK

1. Anstis, S. *Expl Brain Res.* **103**, 476–478 (1995).
2. Rieser, J. J., Pick, H. L., Ashmead, D. H. & Garing, A. E. *J. exp. Psychol., hum. Percept. Perform.* **21**, 480–497 (1995).
3. von Helmholtz, H. *Physiological Optics* Vol. III (Dover, New York, 1962) (tr. from 3rd German edn of 1910).
4. Barlow, H.B. in *Vision: Coding and Efficiency* (ed. Blakemore, C.) 363–375 (Cambridge Univ. Press, 1990).

* Also at: Visual Development Unit, University College, Gower Street, London WC1E 6BT, UK.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.

Learning improved by arts training

SIR — It has been suggested^{1,2} that musical experience or training can temporarily strengthen visuospatial reasoning. We report here further data on the relationship between arts training and some broader aspects of learning.

In its first year, our study included 96 pupils (aged 5–7 years) in eight first-grade public-school classrooms. Four 'test arts' classrooms (two each in two schools) participated in a music and visual-arts curriculum which emphasized sequenced skill development³. Two other control, 'standard arts' classrooms in each school received the system's standard visual arts and musical training. Curricula in all classrooms were otherwise identical.

After seven months, the schools gave all students standardized First-Grade Metropolitan Achievement Tests, and in the 83% of pupils for whom the details

were already on file, we compared the results with kindergarten achievement scores. Classroom teachers also evaluated each student on attitude and behaviour (on Likert-scale questionnaires) four times during the study.

We found that, in the 83% of students with kindergarten scores on file, those in the test arts classes started behind the control children in having kindergarten tests at least at the national average grade level (38% compared with 47%; $P < 0.05$, χ^2), but after seven months, they had caught up to statistical equality on reading and were now ahead on learning mathematics (77% compared with 55% were now at grade level or above; $P < 0.05$, χ^2).

In the full first-grade sample, pupils in the test group were again equal to those in the control group on reading, but ahead on maths (75% compared with 53% at or above grade level; $P < 0.05$, χ^2). This maths advantage did not depend on school, or on whether students entered after poor, average or good kindergarten performance (see figure). The results of the

achievement tests were also above grade level: in the full sample, for example, the percentage increment of test arts pupils over controls fell only gradually over levels above the 50th percentile, and were still present to a small degree even at the 90th percentile, whereas in reading, the percentages between the groups remained roughly equal over higher percentiles.

Classroom attitude and behaviour ratings of test students began significantly behind those of controls, possibly reflecting their poorer kindergarten start, but reached statistical equality by achievement testing time. This might explain the equality between the groups on reading, but not fully the improvement in maths.

We continued the study the following year in four test and five control second-grade classrooms at the same schools, all students again taking achievement tests after seven months. We found that test and control arts pupils were again equal on reading, and that test arts pupils were again ahead on maths (73% compared with 55% at or above grade level on comprehension, $P < 0.05$, χ^2 ; 71% compared with 63% on problem-solving). The percentage of students at or above grade level in second-grade maths was highest in those with two years of test arts, less in those with only one year and lowest in those with no test arts (on maths comprehension; $P < 0.01$, CMH statistics).

We believe our data show that when students discover that participation in arts activities is pleasurable, they become motivated to acquire skills in the arts on which our programme focuses, with two types of result. First, from realizing that they can learn such challenging but desirable skills, students' general attitude towards learning and school can improve. Second, learning arts skills forces mental 'stretching' useful to other areas of learning: the maths learning advantage in our data could, for example, reflect the development of mental skills such as ordering, and other elements of thinking on which mathematical learning at this age also depends⁴.

Martin F. Gardiner

Alan Fox

*Research and Outreach Programs,
The Music School,
Providence, Rhode Island 02906, USA*

Faith Knowles

*Kodaly Center of America,
Providence, Rhode Island 02906, USA*

Donna Jeffrey

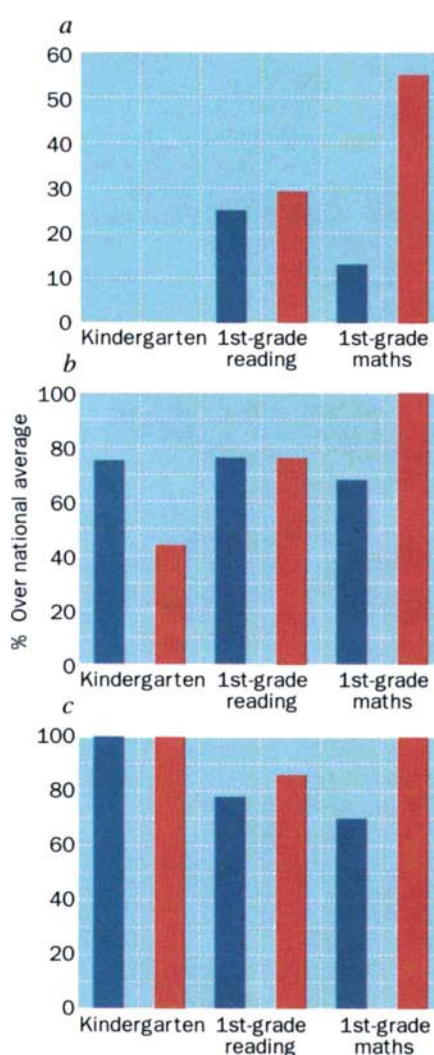
*Start With Arts Program,
Pawtucket, Rhode Island 02860, USA*

1. Rauscher, F. H., Shaw, G. L. & Ky, K. N. *Nature* **365**, 611 (1993).

2. Rauscher, F. H., Levine, L. J., Wright, E. L., Shaw, G. L. & Ky, K. N. Paper presented at 102nd a. Convention Am. psychol. Ass. (Am. Psychol. Ass., Los Angeles, 1994).

3. Chosky, L. *The Kodaly Concept: Creating an Environment for Musical Learning* (Prentice-Hall, Englewood Cliffs, NJ, 1981).

4. Piaget, J., Szeminska, A. & Hodgson, M. *The Child's Conception of Number* (Routledge & Kegan Paul, London, 1952) (tr. Gattengo, C.).



First-grade achievement test performance by previous kindergarten performance. a, Scores below 34th national percentile; kindergarten scores not available. b, Kindergarten scores between 34th and 66th national percentile. c, Kindergarten scores above 66th national percentile. Blue bars, standard arts; red bars, test arts.

METHODS. The 80 students in the first-grade sample for whom kindergarten achievement test scores were also available were divided into three subgroups, a, b or c, as defined above. For each subgroup, the per cent of test and control arts students at national average (grade level) or above (that is, at 50th national percentile or above) were calculated and compared for kindergarten, first-grade reading and first-grade maths achievement tests. In a, some students could raise performance to grade level or above in reading by the end of the first grade, roughly 25% in control and test arts groups, but in maths a significantly higher percentage of test arts students do so ($P < 0.05$, χ^2). In b, although more test students are below grade average at kindergarten level, after first grade they are equal to controls on reading and ahead on maths ($P < 0.06$, χ^2). Even with the best kindergarten test students (c), there is improved learning in maths in test arts students ($P < 0.03$, χ^2). Such uniformity of the test versus control difference across strata adds to the statistical significance of test versus control difference in the full combined strata ($P < 0.01$, CMH statistics). (In the classes studied, students receiving remedial teaching for limited proficiency in English, and those entering from a transitional year, repeating the first grade or not present for the full year, were analysed separately, and are not included in the present results.) Full details of methodology and other results described here are available directly from the authors (mgardi1140@aol.com).

Quest for low-degree mantle melts

SIR — Baker *et al.*¹ report an experimental study at pressures of 1 GPa (1,000 million Pa) to determine the composition of a silicate melt produced at anhydrous, near-solidus conditions from a lherzolite composition. They obtained a composition at 1,250 °C with ~57% SiO₂ and ~5.7% Na₂O, which is 'andesitic' in the sense of its low normative diopside (5%), high normative hypersthene (>15%) and high normative plagioclase (>75%, andesine). These results conflict with earlier conclusions that anhydrous melts of fertile lherzolite at 1 GPa are basaltic, with ~15–20% normative diopside, >10% normative olivine and, at low degrees of melting, nepheline-normative². From our own experience and experimental tests, we suggest the following possible explanations for the observations of Baker *et al.*

(1) The diamond-aggregate technique they used retains internal pore pressure less than the experimental load pressure, until the pores are filled by the melt phase. Initial melt compositions may thus reflect $P \ll P_{\text{load}}$, and may tend towards the 1-atmosphere minimum melt composition (which is quartz-normative). Adjustment to higher pressure equilibrium is diffusion-controlled in relation to the lherzolite mineralogy outside the diamond aggregate. Tests of the diamond-aggregate technique at higher degrees of partial melting and higher temperatures (>1,300 °C) are not necessarily transferable to low-degree melting experiments at 1,250 °C.

(2) In Baker *et al.*'s 'two-stage experi-

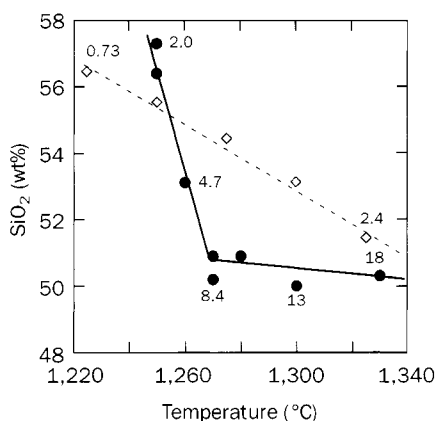


FIG. 1 SiO₂ contents versus temperature for peridotite partial melts determined using the diamond-aggregate technique for the peridotite composition MM3^{1,5} compared with peridotite partial melts from MM3 modelled using the MELTS program¹. Note the linear variation produced by MELTS versus the 'dog-leg' trend of the diamond-aggregate experiments. The number next to data points is the measured or calculated degree of melting; note the large discrepancy between MELTS and the experimental results on MM3^{1,5}.

ments' the melt composition migrating into the diamond aggregate could be a remobilized, quench-modified melt from the stage-one experiment; that is, a remelted glass derived by olivine + quench clinopyroxene growth during quenching of the melt formed from the mineral mix in the stage-one experiment.

(3) The use of a mix of naturally occurring olivine, aluminous pyroxenes and Cr–Al spinel mineral separates from spinel lherzolites as starting materials introduces considerable uncertainty in distinguishing between disequilibrium and equilibrium melting. Aluminous clinopyroxene and orthopyroxene, formed at higher pressures and probably lower temperatures, have been shown to melt incongruently and to exsolve a SiO₂-, Na₂O- and Al₂O₃-rich glass when held at lower pressures and temperatures of 1,100–1,300 °C (ref. 3).

Disequilibrium melting of a mineral assemblage run outside its pressure and temperature stability field can produce a liquid at temperatures below the true solidus. The methodology of the mineral mix and diamond-aggregate technique requires crushing of the diamond aggregate, and identification and analysis of juxtaposed residual phases adequately, nor to identify whether glass formed within phases or at particular mineral interfaces.

(4) The movement of melt into pore spaces of the diamond aggregate is a time-dependent process, as is the reaction, melting and attempted re-equilibration of the natural mineral mix. The time-dependent separation of melt from its reactants is obviously undesirable, and diffusion-controlled run products must be expected. Baker *et al.* note clinopyroxene crystallization within the diamond aggregate and mineral zoning in the peridotite. In these circumstances, the 'reacting composition' is unknown and is not the same as the bulk peridotite composition.

(5) Baker *et al.* initially considered that their experiments were anhydrous, but in their correspondence with us they argue that we have not incorporated the 'right amount' of water to reproduce their results in our reversal experiments (our experimental glasses are H₂O-free, as demonstrated by infrared spectroscopy). Unknown and very high water contents in melts of lherzolite at 1 GPa will indeed produce quartz-normative liquids, albeit of a different nature from that found by Baker *et al.*⁴.

The experimental melting data for MM3 is compared with liquids calculated by the MELTS program¹ in Fig. 1. Figure 2 illustrates the disagreement between

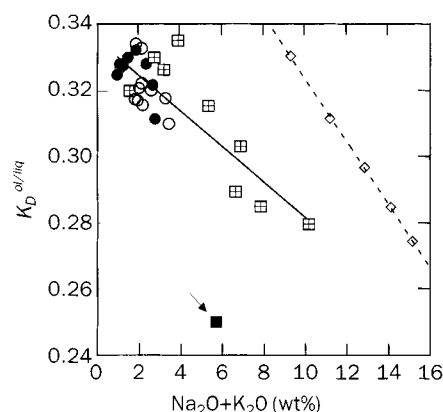


FIG. 2 Olivine/liquid Fe/Mg exchange coefficient (K_D) determined for the Baker *et al.*¹ result compared with other experimental data over a large range of alkali contents and with values calculated by MELTS¹. ■, Run 70a in ref. 1 (arrowed); ○, peridotite sandwich experiments of ref. 2; ●, diamond-aggregate experiments of ref. 5 at temperatures >1,270 °C; □, olivine/liquid K_D determined in ref. 6 in the system SiO₂–K₂O–FeO–MgO; ◇, modelling using MELTS at QFM for MM3 (ref. 1) at 1 GPa.

olivine/liquid partitioning of Fe and Mg as measured by E. Takahashi and calculated by MELTS, and the 1,250 °C data point published by Baker *et al.*¹. Both figures presented here suggest disequilibrium melting and quenching problems in the recent experiments.

To summarize, the experimental results of Baker *et al.*¹ must be treated with caution. The experimental problems are twofold, relating first to time-dependent pressure gradients, time-dependent diffusion and time-dependent pore filling of the diamond aggregates; and second, to disequilibrium melting of natural mineral starting mixes. A third problem — uncertainty in water content — is possibly an additional source of error.

Trevor J. Falloon*

David H. Green

Hugh St C. O'Neill

Research School of

Earth Sciences,

The Australian National University,

Canberra ACT 0200, Australia

and

**Geology Department,*

University of Tasmania,

GPO Box 252C,

Hobart, Tasmania 7001,

Australia

Chris G. Ballhaus

Max-Planck-Institut für Chemie,

Postfach 3060,

Mainz 55128, Germany

1. Baker, M. B., Hirschmann, M. M., Ghiorso, M. S. & Stolper, E. M. *Nature* **375**, 308–311 (1995).

2. Falloon, T. J. & Green, D. H. *Miner. Petrol.* **37**, 181–219 (1987).

3. Doukhan, N., Doukhan, J.-C., Ingrin, J., Jaoul, O. & Rateron, P. *Am. Miner.* **78**, 1246–1256 (1993).

4. Green, D. H. *Can. Miner.* **14**, 255–268 (1976).

5. Baker, M. B. & Stolper, E. M. *Geochim. cosmochim. Acta* **58**, 2811–2827 (1994).

6. Takahashi, E. *Geochim. cosmochim. Acta* **42**, 1829–1844 (1978).

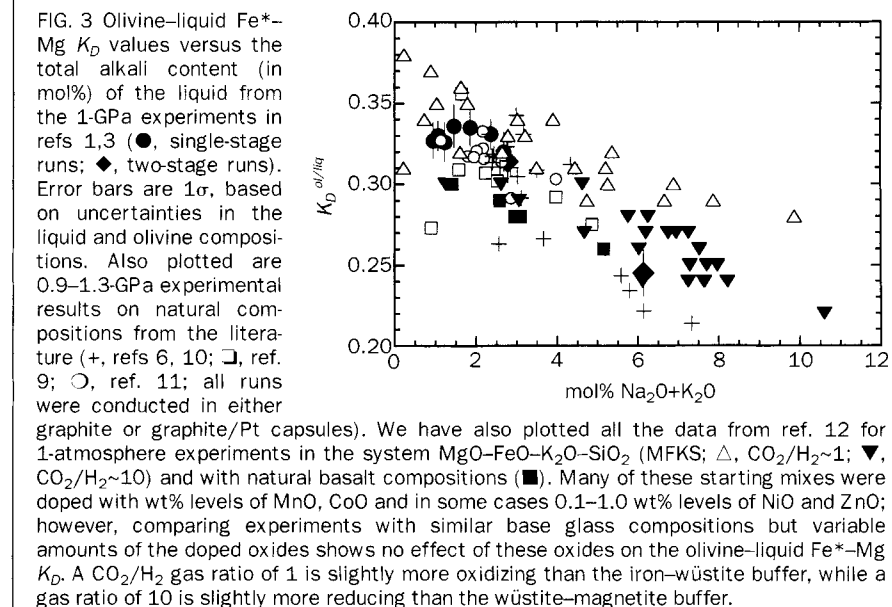
BAKER *ET AL.* REPLY — Falloon *et al.* assert that the melts generated in our near-solidus two-stage diamond-aggregate experiments¹ represent disequilibrium liquids. Although space limitations prevent us from describing our current experimental work, we address each of their points below.

(1) At the onset of a single-stage diamond-aggregate experiment, low pore pressure in the diamond layer could affect the composition of liquid initially filling these pores². Our two-stage diamond-aggregate experiments were designed to overcome this problem, as the first liquid mobilized in the second stage is expected to be the remobilized 1-GPa-equilibrated melt from the first stage of the experiment. The small amount of clinopyroxene (cpx) in the glass from the diamond layer and the residual cpx in the peridotite layer at 1,250 °C¹ have the same compositions — analyses of all oxides overlap at the 1–2 σ level (analytical uncertainties are similar to those reported for cpx in ref. 3). This similarity in cpx compositions provides strong evidence for a close approach to equilibrium between melt in the diamond layer and residual cpx in the peridotite layer.

(2) Single- and two-stage experiments at 1,270 °C yield nearly identical melt compositions³; therefore, for two-stage runs, any quench modification at the end of the first stage is erased during the second stage. Further, mass-balance calculations show that liquids from our 1,250 °C experiments cannot be produced from more 'typical' melts of peridotite (for example, the 1,270 °C liquids) by crystallization of any combination of quench cpx and olivine.

(3) The initial melting of natural minerals or synthetic oxides is nearly always expected to be a disequilibrium process — the important point is whether the liquid subsequently approaches equilibrium with the residual solid phases over the course of the experiment. To our knowledge, pyroxene premelting phenomena^{4,5} occur only within single phases and are unlikely to affect melt formation along multiphase grain boundaries. Even if such melts did form on grain boundaries, it is difficult to understand how they could persist in intimate contact with the spinel-lherzolite assemblage on the timescales of the first stages of our two-stage experiments.

Falloon *et al.* plot the results of MELTS calculations on MM3 against temperature. We have not suggested that silica-rich partial melts of peridotite form through low temperature, but rather that they form at low melt fractions when the alkali content of the melt is high, and in this regard the experiments and calculations are in excellent agreement (see Fig. 2a of our paper¹).



Although they agree with our suggestion from our data¹ that olivine-liquid Fe*-Mg K_D decreases with increasing liquid-alkali content (Fe* denotes all Fe as FeO), Falloon *et al.* use data selectively to conclude that our results are anomalous. Figure 3 here shows a more complete data set on olivine-liquid Fe*-Mg K_D versus liquid-alkali content for both simple and complex melts, and it is clear that our 1,250 °C K_D value is within the range of experimental values. Olivine-liquid K_D s calculated from 16–20-kbar experimental data^{6,7} also show low values (<0.26) for liquids with total alkali contents of >5 mol%.

(4) As we stated in point (1) above, two-stage experiments greatly minimize the need for diffusion-controlled re-equilibration in the second stage. The presence of small amounts of cpx in the diamond layers of some of our experiments is unlikely to have affected the 'reacting composition', because the compositions of these pyroxenes are similar to those in the residual peridotite. Similarly, zoning in residual cpx cannot have a significant effect on partial melt compositions because core and rim compositions overlap at the 1–2 σ level for all oxides (analytical uncertainties are similar to those reported for cpx in ref. 3).

(5) Nowhere have we claimed that our experimentally produced glasses are anhydrous. Small amounts of water are present even in nominally anhydrous peridotite⁸, and this water can become concentrated in melt at low melt fractions. We therefore deliberately used the expression 'nominally anhydrous' when describing our results (page 311 of our paper). However, although it is likely that our lowest-temperature experimental glasses contain 0.5–1.0 wt% H₂O, this is insufficient to have significantly influ-

enced the trends observed in the glass compositions in the lowest-temperature experiments.

In summary, the various tests of our technique and comparisons with the MELTS calculations we have presented previously support our view that the compositional trends displayed by all our 10-kbar partial melts^{1,3} are correct.

M. B. Baker

M. M. Hirschmann*

L. E. Waslylenki

E. M. Stolper

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

M. S. Ghiorso

Department of Geological Sciences, University of Washington, Seattle, Washington 98195, USA

1. Baker, M. B., Hirschmann, M. M., Ghiorso, M. S. & Stolper, E. M. *Nature* **375**, 308–311 (1995).
2. Johnson, K. T. & Kushiro, I. *Geophys. Res. Lett.* **19**, 1703–1706 (1992).
3. Baker, M. B. & Stolper, E. M. *Geochim. cosmochim. Acta* **58**, 2811–2827 (1994).
4. Doukhan, N., Doukhan, J.-C., Ingrin, J., Jaoul, O. & Rateron, P. *Am. Miner.* **78**, 1246–1256 (1993).
5. Rateron, P., Ingrin, J., Jaoul, O., Doukhan, N. & Elie, F. *Phys. Earth planet. Inter.* **89**, 77–88 (1995).
6. Kinzler, R. J. & Grove, T. L. *J. geophys. Res.* **97**, 6885–6906 (1992).
7. Takahashi, E. & Kushiro, I. *Am. Miner.* **68**, 859–879 (1983).
8. Bell, D. R. & Rossman, G. R. *Science* **255**, 1391–1397 (1992).
9. Hirose, K. & Kushiro, I. *Earth planet. Sci. Lett.* **114**, 477–489 (1993).
10. Bartels, K. S., Kinzler, R. J. & Grove, T. L. *Contr. Miner. Petrol.* **108**, 253–270 (1991).
11. Falloon, T. J. & Green, D. H. *Miner. Petrol.* **37**, 181–219 (1987).
12. Takahashi, E. *Geochim. cosmochim. Acta* **42**, 1829–1844 (1978).

* Also at: Department of Geology, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

Evolutionary vistas

Mark Ridley

Narrow Roads of Gene Land: The Collected Papers of W. D. Hamilton. Vol. 1 — Evolution of Social Behaviour. By W. D. Hamilton. W. H. Freeman/Spektrum: 1996. Pp. 552. £40 (hbk); £20, \$25.95 (pbk).

W. D. HAMILTON is a giant of evolutionary biology who has brought about two major revolutions. The first concerned altruism. Natural selection, it might seem, should theoretically work against altruistic, or cooperative, behaviour in which one individual reduces its reproduction for the sake of another. Altruistic behaviour nevertheless happens, and in 1964 Hamilton showed how natural selection can favour it; his explanation has far-reaching consequences and is generally accepted to be correct. The second revolution was sexual. Natural selection, again theoretically, should not favour sex, and the existence of sex is the outstanding puzzle in evolutionary biology. Since 1980 Hamilton has been writing more about it, suggesting that the advantage of sex lies in the relationship between hosts and parasites. He has redirected research, but the jury is still out on whether his explanation is correct. In addition to these two important contributions, Hamilton has published classic papers on many related topics in the evolution of sex and social behaviour.

Narrow Roads of Gene Land is the first volume of his papers and contains his 15 main publications from 1963 to 1980. They are, as the subtitle says, mainly about social behaviour. A second volume will contain subsequent publications, mainly about sex. Volume 1 has the classic papers on altruism, senescence, sex ratio, fig wasps, insects in rotting wood, and dispersal. They range (to develop Hamilton's own titular metaphor) from the 1980 paper on dispersal, a cooperatively constructed and little-known route that Hamilton himself is not sure he understands, and that may lead nowhere, or into a private territory where only Australian equation writers can survive; through immediately popular interstate highways such as his 1967 work on sex ratios; to the unique mountain-pass that is his 1964 paper on altruism — forged by himself in an almost private project that the few who did know about it thought would lead only into sterile or political no-go zones. The dynamite blasts sometimes failed to ignite, or blew back, but he eventually cleared the way into what later grew into some of the most fer-

tile land in evolutionary biology.

The significance of the 1964 paper is no longer doubted (except by a few politico-paths); I guess it is among the 20 or so most important papers on the subject. The reprinting of the papers is valuable in itself. Some readers will visit them for the first time. Others will retrace their foot-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

W. D. Hamilton: imagines the creatures around him in almost mythic relation with himself.

steps, and not regret it; as Hamilton says: "one reads and forgets". His papers contain (as well as their well-known main themes) myriad incidental remarks on little-studied insects and unexplored theoretical problems, and these could do more to rejuvenate the subject than several conference-loads of grey 'new approaches'. Hamilton is always a creative thinker and rarely touches a subject without offering an original thought.

The papers themselves are the core of the book; but it will, I suspect, be the autobiographical introductions that people will initially buy the book for and read. Hamilton has written a nontechnical introduction of 10 pages or so to each

paper, and together they make up almost 150 pages of this volume. Perhaps there will be a similar amount in Volume 2, where I hope scholarly gravitas will be relaxed to admit Hamilton's extraordinary and sometimes astonishing essays from the *Times Literary Supplement* about his early experiences of natural history. The whole will then effectively add up to an autobiography: a significant fact, because it turns out that Hamilton is something of a wizard at the genre.

His autobiographical virtues sometime seem to me to be of a piece with his research. There is the same originality and honesty, together with a strong reaction to his natural surroundings. He likes to imagine the creatures around him in almost mythic relation with himself. I can do no more than mention some of the themes. There is good stuff on London, and the enigmatic George Price who found him there; on first conferences (baffled by the luxury, including the refrigerator bar, and cornered by a senator's wife); on Brazil (delightful students and figs, compromised by dangerous coffee); on London's Imperial College (where the undergraduates will surely issue writs of *scandalum magnatum* for his unflattering observations, to follow their earlier petitions about his teaching); and on libraries. My own modest career has been in many respects the opposite of Hamilton's, but I am pleased to find one shared experience. I too have failed to penetrate only one library, and like him it was that "very private and patrician" institution, the Zoological Society of London.

All this Hamilton relates in a unique style; I cannot place it among any of the standard models, but you soon get into it. The papers in the book are already classics of scientific research, and the introductions deserve to become classics of scientific autobiography. The scientist's mind they reveal, however, is a highly idiosyncratic one, and anyone who thinks there is any such thing as 'the scientist' should read Hamilton, and J. D. Watson, and think again.

So what was it like, to work out the theory of social behaviour? Hamilton was ostensibly a graduate student at the time, in London. The intellectual support was not good. Supervisors (mainly in the literal sense, of overlook), genetic and non-genetic alike, worried about the genes for altruism ("could genes really affect altruism, he wondered, was there any evidence?"), doubted there was a problem at all (it was a fact the whole world knows, that natural selection guaranteed the good of the species) or detected "a sinis-

James King-Holmes

ter new sucker budding from the roots of the recently felled tree of Fascism". The institutional support was no better. Hamilton's department did not even supply him with a desk, and he did the work in a room in Chiswick or on the platform at Waterloo Station. *Nature*, too, played its part. The short 1963 paper that provides the first statement of his theory of altruism was submitted there. "I received the editor's decision almost by return of post. In about three lines he regretted [...etc.] and suggested that... it might be more appropriate in a 'psychological or sociological [sic] journal'."

Science is a most inequalitarian exercise, as a tiny minority of research accounts for an extravagant amount of the permanent value in the subject. Thirty-five years on, the forces of conformism are as powerful as ever and it is easy to predict that the next Hamilton will also do his work in an obscure garret. But however that may be, the rest of us can meanwhile enjoy the collected papers of our own Hamilton, papers that contain more of value than a shelf of the officially sponsored 'mainstream' work that (as he recalls) he never did manage to immerse himself in. □

Mark Ridley is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

New Journals

This year, *Nature's* annual New Journals review supplement will appear in the issue of 5 September. Publishers and learned societies are invited to submit journals for review, taking note of the following criteria:

- Journals that first appeared during or after June 1994 and issued at least four separate numbers by the end of April 1996 will be considered.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are publications of abstracts.
- Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.
- Deadline for submission is 14 June.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates. For further information please contact Peter Tallack, *Nature*, Macmillan Magazines Ltd, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567; e-mail: p.tallack@nature.com.

Mind maps

Roy Porter

The History of Mental Symptoms: Descriptive Psychopathology since the Nineteenth Century. By German E. Berrios. Cambridge University Press: 1996. Pp. 565. £100, \$165 (hbk); £40, \$59.95 (pbk).

Is madness real? Anti-psychiatrists such as Thomas Szasz, who wrote an influential work entitled *The Myth of Mental Illness*, have notoriously denied the reality of psychiatric disorders. But to many outsiders this is not psychiatry's Achilles' heel: few, after all, dispute that deeply depressed people are sick or that treatments help. What seems far more worrying is that the diagnoses psychiatrists offer seem to come and go; new terms appear on the diagnostic charts and old ones drop off like changing Paris fashions. A century ago, neurasthenia was the flavour of the month, but no one is neurasthenic nowadays. And what about hysteria? It made Freud's name, but try looking for it in the most recent *Diagnostic and Statistical Manual* of the American Psychiatric Association, although you will find such syndromes as 'female sexual arousal disorder', which suggests that our diagnostics reveal more about the hang-ups of Western society and the psychiatric profession than about disease itself.

So what do psychiatric diagnostics amount to? Are they fact or factitious — fictions even? German Berrios's history of descriptive psychopathology, a *tour de force* of multilingual erudition, surveys the sources, provides a perceptive analysis and offers some broad reflections on the Herculean enterprise of mapping a sea of troubles. Well versed in English, French, German, Italian and Spanish sources, Berrios, who teaches psychiatry at Cambridge, England, divides up the field thematically (examining anxiety, obsessions, personality disorders and so on) and then confidently pilots us through the intricate nosological charts, the matchings, rematchings and mismatchings of symptom clusters and disease labels successively proposed from 1800 to the present. The organic substratum of mental illness may itself undergo secular change, he reflects, but his study aims to plot the

shifting concepts of the discipline itself, and the mutable meanings of stable terms. ('Depression' does not mean the same now as in 1796.)

The overall picture is one of a real long-term consolidation of the terminology of descriptive psychopathology. But Berrios does not Whiggishly conclude that this is proof of 'progress', an ascent from error to science. Rather, he stresses factors favouring consensus, how the concentration of vast patient populations in nineteenth-century asylums gave the emergent psychiatric profession opportunities for meticulous study of symptoms, especially over protracted time spans. Professional pressures also played their



"Autoportrait" by Herbert Bayer, 1932. From the cover of *Petit traité de médecine psychosomatique* by Jocelyne Vaysse. Synthélabo, FF84 (pbk).

part: in nineteenth-century France, Esquirol ruled the specialty and so his taxonomy assumed dominance; the same happened with Kraepelin in Germany; and the *Diagnostic and Statistical Manual* is now the authority in the Anglo-American world.

But what specially fascinates Berrios are discrepancies and discontinuities in disease categories. Even today, psychiatric terminology betrays profound national disparities, and linguistic idiosyncrasies have played no small part — no single English word translates *l'âme* (soul, mind, spirit, psyche) or *fou* (mad, foolish), or for that matter Freud's *das Es* (our 'id' seems a curious cop-out). An admirer of the philosopher Gaston Bachelard, Berrios insists that whatever the organic basis of psychopathology, patients don't come with diseases written

on their faces; diagnosis is shaped by the clinician's cognitive frame (Thomas Kuhn's 'paradigms'), and there is no such thing as an atheoretical symptom language: every map has its grid, every picture its frame. Indeed, the psychiatric diagnostician, he suggests, should be seen as a cross between a botanist cataloguing flowers in the garden and a sculptor carving shapes out of marble. Viewed over time, these disease taxonomies reveal remarkable 'ruptures' (Bachelard's term), as is here demonstrated in subtle case studies of delusions, delirium, dementia and a hundred other conditions.

Berrios is far too astute an historian to offer any simple reductionist explanations for shifts in disease classification. He suggests our best explanatory bet lies in the

model of change developed by the great French historian Fernand Braudel, with its harmonic interplay of short-, medium- and long-term forces. New theoretical schemes have always been important, but so have practical factors, for example the opportunities for the study of symptoms that were afforded by the invention of photography.

Berrios sees his book as a beginning. For the meantime, he has put us deeply in his debt with a remarkable account of the mappings of the mind through a study that transcends the private technicalities of psychiatry to shed light on the changing representations of the Western psyche itself. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Pheromones, drugs and exploding bugs

Tristram D. Wyatt

Bombardier Beetles and Fever Trees: A Close-Up Look at Chemical Warfare and Signals in Animals and Plants. By William Agosta. Addison-Wesley: 1996. Pp. 224. \$25.

Chemical Ecology: The Chemistry of Biotic Interaction. By Thomas Eisner and Jerrold Meinwald. National Academy Press: 1995. Pp. 214. \$49.95, £34.95.

POISON arrow frogs, the arms race between plants and their insect herbivores, and lobsters calling for mates are all examples of chemical defences or signals. They form part of a hidden chemical world of exquisite variety and subtlety only recently revealed to us.

Chemical interactions between organisms are explored in these two new books. William Agosta's is written for lay people whereas Thomas Eisner and Jerrold Meinwald's intended audience is the wide scientific community, from student to researcher. Both books offer a showcase for chemical ecology, a subject coming of age as advances in chemistry are combined with more sophisticated ecological, as well as molecular, studies. A good case is made by both books for the intellectual excitement of chemical ecology, but both also focus on its great practical uses in, say, drug discovery — cyclosporin and ivermectin being recent examples.

Eisner and Meinwald bring together a fine collection of chapters from a symposium on chemical ecology. The subjects range from chemical defences used by animal and plants, pheromones in ant colony organization, algal pheromones, sexual selection based on chemicals, eavesdropping and deceit by predators and parasites, through to the production and reception of pheromones and other signals at the

molecular level. It is deliberately not an exhaustive survey: most chapters start with an overview and then focus on a particularly well-studied model system, such as aquatic chemical signals which are beautifully illustrated by lobsters, signal transduction by activation of helper T cells, and nervous-system processing of signals by the moth brain/pheromone system. (The lack of a chapter on mammal or other vertebrate pheromone systems is a shame.)

Although some chapters contain familiar stories, all offer first-rate accounts of recent advances and give excellent examples. Most chapters are lively and there are some useful photographs. Surprisingly, there are no links between chapters. The strength of the book is its range, from ecology and evolution to molecular mechanisms. It succeeds as an exciting introduction to chemical ecology.

Agosta's subtitle is "chemical warfare and signals in animals and plants", but his book, although touching on many examples covered in Eisner and Meinwald's volume, includes much more than this. His theme is chemical adaptation in the broadest evolutionary sense. Structural chemical adaptations of spider silk and deer antlers are found alongside the chemical duets of fungus hyphae and the more familiar moth sex pheromones. The chemical adaptations of ice-fish and mites for coping in Antarctica, the production of light in fireflies and the use of pigments in animal colours are also covered. There is

inevitably some overlap of material with the author's well-received *Chemical Communication* (W. H. Freeman/Scientific American Library, 1992), but this is small. The examples are up to date and the text is well edited. Two things that would improve the next edition are suggestions for further reading (or references — possible and useful even in popular books) and better illustrations.

Agosta is adept at handling controversial topics such as the existence of human pheromones for attraction between the sexes, and does not present them as cut and dried. He is also good at linking topics to their human interest. For example, he addresses environmental issues when comparing pesticides with the alternatives offered by microbial insecticides or by mating disruption with pheromones, and looks at the wider issues of drugs such as quinine and morphine. Particularly nicely done is his description of chemical cues in the life cycle of two parasites that affect millions of people: the tropical plant parasite *Striga*, or witchweed, with its minefield of seeds ready to germinate when crops are planted; and the complex life history of the schistosomiasis parasite, from snail to people.

Agosta's book is a lucid account of the chemistry behind the natural world. As



Undercover protection: a female Australian sawfly (*Pseudoperga guerini*) guarding a clutch of larvae recently emerged from her eggs.

part of the current renaissance in popular science books written by leading scientists, it may help to do for chemistry what others have done for evolution and genetics.

Both books end with the theme of biodiversity and its value — for drug discovery and much more — and the devastating threats to it from species extinction, especially through tropical deforestation. It seems likely that most species will be lost even before they have been described, let alone investigated for their chemical interest. Chemical ecologists can provide us with powerful arguments for species conservation. □

Tristram D. Wyatt is in the Departments of Continuing Education and Zoology, University of Oxford, Oxford OX1 2JA, UK.

Essentials extracted

Alexander N. Glazer

Bioconjugate Techniques. By Greg T. Hermanson. Academic: 1996. Pp. 785. \$99 (hbk); \$49.95 (pbk).

MANY unrelated reasons prompt the chemical modification of peptides, proteins, oligonucleotides, sugars and polysaccharides, lipids and glycolipids. The unifying elements are found in the chemical methods most commonly used and in the preferred (or most readily available) reagents. This book deals with the chemistry of the reactive groups on target molecules; the reagents for crosslinking or tagging molecules; and the many unrelated applications of conjugated molecules — in immunology, as affinity supports, in liposome technology, in the

preparation of proteins labelled with colloidal gold, as enzyme conjugates, in the labelling of nucleic acids, in enzyme conjugation to DNA and so on.

The presentation throughout is that of a laboratory manual. Well-illustrated descriptions of a particular chemical reaction are followed by a detailed protocol and list of references extending into 1994. The coverage of reagents and methods is comprehensive, and sources are given for commercially available compounds. The breadth of coverage inevitably leads to superficial discussion of individual topics, so readers would be well advised also to consult authoritative primary sources. Typographical mistakes, occasional factual errors and stylistic inconsistencies in the reference list mar an otherwise well-produced and useful text. □

Alexander N. Glazer is in the Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3206, USA.

New in paperback

Darwin's Dangerous Idea: Evolution and the Meanings of Life by Daniel C. Dennett. Touchstone, \$16. "Never in the field of scientific endeavour can so great a theory have been so misunderstood by so many with so little reason: but Dan Dennett's book is a marvellous corrective", Mark Ridley, *Nature* **375**, 457 (1995).

The Consumer's Good Chemical Guide: Separating Facts from Fiction in Everyday Life by John Emsley. Corgi, £6.99. Winner of the 1995 Rhône-Poulenc Prize for Science Books. "Accessible and entertaining... deserves to be widely read", John Mann, *Nature* **371**, 214 (1994).

About Time: Einstein's Unfinished Revolution by Paul Davies. Penguin, £8.99. "Almost every possible aspect of 'time' and its associated scientific and philosophical problems are touched on in this book... [it] can be read easily and most readers will find in it matters of absorbing interest", Peter Landsberg, *Nature* **375**, 548 (1995).

Evolution and Healing: The New Science of Darwinian Medicine by Randolph M. Nesse and George C. Williams. Phoenix, £7.99. "A needed preventive measure against the enduring, endemic nineteenth-century consciousness about nature that pervades medical thinking near the turn of the twenty-first century — a vaccine, one might say, against the lulling delusions of pre-Darwinian thought... [the authors] give both physicians and lay readers a chance to remedy, quite painlessly, the ignorance of evolution that expensive educations have usually left them with", Melvin Konner, *Nature* **375**, 641 (1995).

The Third Culture: Beyond the Scientific Revolution by John Brockman. Touchstone,

\$14. "Brockman has assembled two-dozen voluble scientists, most of them with a commitment to popular exposition, and in interviews, from which the editor's prompts have been expunged, gives them their head... mainly the tributes go something like this: 'Jack is one of my most cherished friends. No-one can have a higher opinion of him than I and I think he is an unctuous, boneheaded little creep'", Walter Gratzer, *Nature* **375**, 743 (1995).

Marie Curie: A Life by Susan Quinn. Addison-Wesley, \$16. "Quinn has produced a magisterial piece of work... wonderful in its depth and detail, so much so that it will be years, if ever, before this account is displaced", June Goodfield, *Nature* **374**, 831 (1995).

The Coming Plague by Laurie Garrett. Penguin, £12.50. Winner of the 1996 Pulitzer prize, this is an impressively encyclopaedic account of the threat posed by newly emerging viruses.

A Moment on the Earth: The Coming Age of Environmental Optimism by Gregg Easterbrook. Penguin, \$14.95. The author of this controversial tome contends that most Western environmental trends are improving and suggests that we should adopt a 'middle path' — an approach that recognizes the seriousness of current human abuses of the environment but also the strength and enduring power of nature.

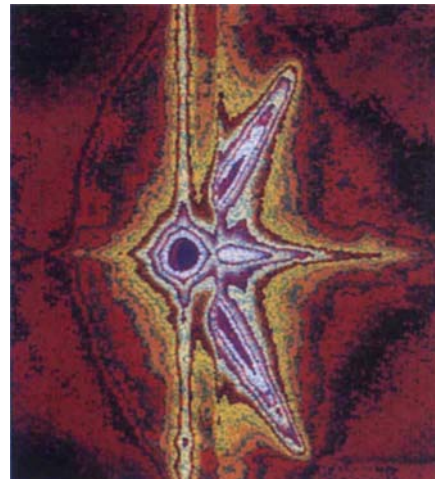
Selected Journals and Other Writings by John James Audubon, edited by Ben Forkner. Penguin, \$15.95. Audubon is famous for his vivid paintings of American birds, several of which are reproduced here. But this volume is mainly a showcase for his lesser-known literary talent: journals, letters, bird biographies, a memoir of his early years and stories of the American frontier.

Not just saying no

Philip St J. Russell

Photonic Crystals: Molding the Flow of Light. By John D. Joannopoulos, Robert D. Meade and Joshua N. Winn. Princeton University Press: 1995. Pp. 137. \$35, £30 (hbk).

PHOTONIC crystals provide a good instance of how the intuitive tools developed for one field can lead to a cascade of new possibilities in another. These transparent



Light redirected out of the plane of a two-dimensional photonic lattice. Credit: Paul Gourley at Sandia National Laboratories.

dielectrics, periodically structured on a micron scale, mimic for photons many of the richly varied phenomena seen when electrons interact with semiconductor crystals. Perhaps the most famous is the photonic band gap, which, following its conception by Eli Yablonovitch in the late 1980s, had a difficult birth. Doubts were initially raised about whether it could exist at all (unlike the electronic band gap, it does not appear naturally). Now well established, its ability to 'say no' to photons within a certain colour range has caught the imagination of researchers worldwide.

Photonic crystals affect light in many other counter-intuitive ways. For example, they allow the normally super-fast photons to stand still, potentially enhancing all kinds of light-matter interactions. *Photonic Crystals* is a timely and well-written account of this new field (alas in cgs units), and is full of colourful computer-generated illustrations. That the curves are often rather 'bumpy' (too few computed points) and not always completely accurate (the evanescent field distributions on pages 50 and 52 should not be exactly in step with the layers) does not really detract from the book's value and importance. □

Philip St J. Russell is in the Optoelectronics Group, School of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK.

Discovery of hard X-ray pulsations from the transient source GRO J1744 – 28

Mark H. Finger^{*†}, Danny T. Koh[‡], Robert W. Nelson[‡], Thomas A. Prince[‡], Brian A. Vaughan[‡] & Robert B. Wilson[†]

^{*} University Space Research Association, [†] Space Sciences Laboratory, NASA/Marshall Space Flight Center, ES-84, Huntsville, Alabama 35812, USA

[‡] Space Radiation Laboratory, California Institute of Technology, 220–47, Pasadena, California 91125, USA

AN UNUSUAL astronomical source of hard X-ray bursts, GRO J1744 – 28, was discovered recently^{1,2}. The properties of this source differ markedly from those of other known high-energy burst sources—X-ray bursters, soft γ -ray repeaters and γ -ray bursters—suggesting that it may represent a new type of source. The bursts probably arise from unstable accretion of matter onto a compact object², such as a neutron star, but the nature of the object and the origin of the burst instability have not been revealed by observations of the bursts themselves. Here we report the detection of coherent X-ray pulsations, with a period of 467 milliseconds, from GRO J1744 – 28; these are the first persistent pulsations seen in a bursting X-ray source. These pulses and their timing indicate that the object is a magnetized neutron star, accreting gas from a low-mass companion star. The pulsation rate has been increasing during the period of our observations, indicating that an accretion disk has been formed and that the transfer of matter from the disk is spinning up the neutron star. The source of the instability that leads to the bursts remains unknown.

We discovered pulsations from GRO J1744 – 28 as part of our X-ray pulsar monitoring and detection programme, which uses 20–50 keV data from the Burst and Transient Source Experiment (BATSE)³ on the Compton Gamma-Ray Observatory (CGRO). After discovery of the pulsations we initiated on-board epoch folding. The earliest detection of pulsations from the source occurred on 15 December 1995, 13 days after the first detected burst.

Daily pulse frequencies were obtained by epoch-folding each day's data at a range of frequencies and determining the frequency, ν , that maximized the pulse amplitude. Using a fit to these frequencies, the data were then epoch-folded over ~ 8 hour intervals and a pulse phase obtained for each interval by correlating it with a profile model. Local fits to the pulse phases were used as the basis for multi-day epoch-folding in order to study the pulse profile, while a global fit to the pulse phases was used to investigate the binary system parameters and the torques acting on the neutron star.

Figure 1 shows the 20–40 keV pulse profile obtained with data from 10–16 January 1996, when the mean flux in this band was $(1.6 \pm 0.2) \times 10^{-8} \text{ erg cm}^{-2} \text{ s}^{-1}$. The pulsed fraction ((mean flux – minimum flux)/mean flux) of this nearly sinusoidal profile is $24.6 \pm 0.3\%$. Figure 2 shows the pulse phases relative to a linear model. The signature of the orbit is clearly present, as is a quadratic trend due to the pulsar spin-up (rotation frequency increase). We have fitted the pulse phases with a model that includes a circular orbit and a fifth-order polynomial in pulse emission time. Orbital parameters are given in Table 1. No significant improvement in the fit was obtained using eccentric orbits or a higher-order phase model. Figure 3 shows the estimated spin-up rate of the pulsar, which rises by more than a factor of three during the observations, reaching $1.2 \times 10^{-11} \text{ Hz s}^{-1}$. During the same interval the pulsed flux also increased by about

TABLE 1 Orbital elements

Parameter	Value	Units
P_{orbit}	11.8337 (13)	d
$t_{\pi/2}$	2450079.6552 (18)	JD
$a \sin i$	2.6324 (12)	light s
$f_x(M)$	$1.3638 (17) \times 10^{-4}$	$M_{\odot}$
e	$< 1.1 \times 10^{-3}$	

These orbital elements were obtained by fitting pulse phases. P_{orbit} is the orbital period. At the orbital epoch $t_{\pi/2}$, the mean longitude is 90° . The limit to the eccentricity is at the 90% confidence level.

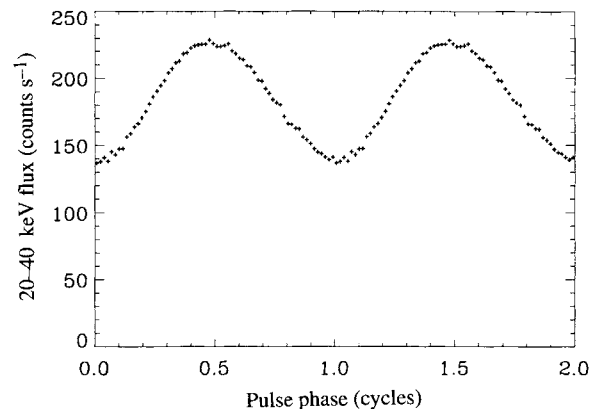


FIG. 1 The pulse profile in the 20–40 keV band for two pulse cycles. This was obtained by combining epoch-folded data and occultation measurements. The data is from BATSE large area detectors 1 and 3 for the time interval 10–16 January 1996. The variable component of the flux is determined from the epoch-folded data while the mean flux is determined from the occultation measurements. Errors are shown for the variation from the mean. There are an additional common error of 2 counts s^{-1} on the mean flux. First and second harmonics of the profile are detected with amplitudes of $6.2 \pm 0.6\%$ and $1.4 \pm 0.6\%$ of the fundamental, respectively. The maximum of the first harmonic occurs 0.164 ± 0.008 cycles before the maximum of the fundamental, indicating a slight asymmetry in the pulse profile.

a factor of three, consistent with the specific angular momentum accreted by the pulsar being roughly constant during these observations. This implies that any counter-torques produced by interactions between the accretion disk and the magnetic field are small compared to the direct accretion of angular momentum.

The pulse profile, spin-up rates and orbital parameters can be used to constrain the inclination angle γ of the neutron star's spin-axis, the angle β between its spin-axis and magnetic pole, its magnetic dipole moment μ , the orbital inclination i , the separation a between the neutron star and its companion, the companion's mass M_c and its Roche lobe radius R_L . This information is essential in understanding the pulsar, its companion, the accretion disk and the burst mechanism.

The highly sinusoidal pulse profile is unusual. Bright accretion-powered X-ray pulsars typically have more complicated pulse shapes⁴. Assuming that the flux modulation arises from a single rotating polar cap which radiates isotropically, the relative pulsed fraction is $m = \tan \gamma \tan \beta$. As the spin axis is probably aligned with orbital angular momentum vector ($\gamma = i$), the observed pulsed fraction constrains the orbital inclination angle, $\tan i = 0.25 \cot \beta$. When gravitational light bending is taken into account, the absence of a peak from a second polar cap restricts $i + \beta < 50^\circ$ for $2GM/Rc^2 = 0.4$ (ref. 5) where G is the gravitational constant, $R \approx 10 \text{ km}$ is the neutron star's radius, and c is the

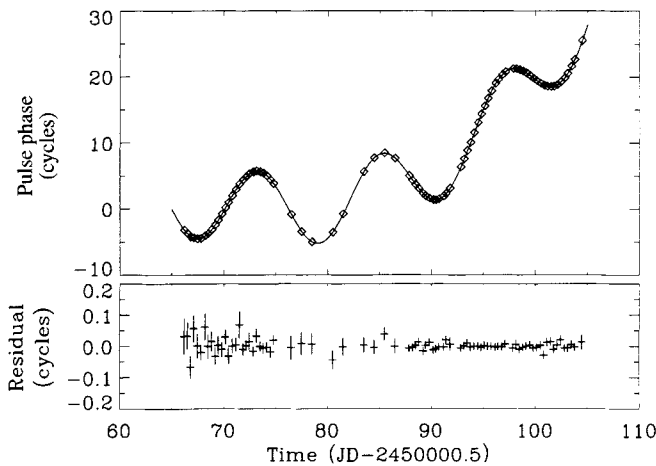


FIG. 2 Top panel, pulse phases relative to a model with a constant barycentric frequency of 2.1409975 Hz. (JD, julian date) The curve is the fit discussed in the text, which includes an orbital model plus a fifth-order polynomial phase model. Bottom panel, residuals to this fit. The data used is from the first pulsed detection on 15 December 1995 (JD 2450066.5) until 23 January 1996. To determine the phases, short (250–300 s) segments of data were fitted to a model that consisted of a Fourier expansion of the pulse profile repeating with the previously determined pulse frequency ν , and a quadratic in time which represents the background. The Fourier coefficients from ~ 8 -hour intervals (1 day for the discriminator rates used between 23 December and 4 January) were then averaged to obtain the Fourier representation of the epoch-folded lightcurve for each interval. In the individual profiles the only significant Fourier coefficient is that at the pulse frequency. The phases were therefore determined for each interval by correlating the profiles.

velocity of light. The observed pulse fraction then requires $i \geq 10^\circ$. Any beaming would reduce this limit.

The X-ray mass function,

$$f_x(M) = M_c^3 \sin^3 i (M_x + M_c)^{-2} = 1.36 \times 10^{-4} M_\odot \quad (1)$$

relates the neutron star and companion masses, M_x and M_c , to the orbital inclination, i ($M_\odot$ is the solar mass). Its unusually small value requires either a low-mass companion or a small orbital inclination (the system being viewed nearly face on). If $M_c \geq 1.0 M_\odot$ and $M_x \geq 1.4 M_\odot$, the inclination must be $< 5.3^\circ$, which for random viewing directions occurs at a probability of 0.43%. Unless strong selection effects favour nearly face-on orbits, the companion star probably has a low mass.

Accretion in this system probably occurs via Roche-lobe overflow. For $M_x = 1.4 M_\odot$, the mass ratio is $q \equiv M_c/M_x \approx 0.05/\sin i$ (for $i > 20^\circ$), and the total binary separation is $a = a_x + a_c \approx 25 R_\odot$. Using the approximate formula of Paczyński⁶, the companion's Roche lobe radius is

$$R_L \approx 0.46aq^{1/3}(1+q)^{-1/3} \approx 4 \sin^{-1/3} i R_\odot \quad (2)$$

where $R_\odot$ is the solar radius. Mass transfer via Roche lobe overflow is thought to begin when a companion with mass $\lesssim 2 M_\odot$ evolves off the main sequence to become a giant with a degenerate helium core and a convective outer envelope⁷. The core mass M_{co} determines the radius of the giant by $R_g \approx 1.3(M_{co}/0.16 M_\odot)^{5/3} R_\odot$. For GRO J1744–28 the companion fills its Roche lobe if $M_{co} \geq 0.21 M_\odot$. The small eccentricity is also consistent with the expected rapid circularization ($t_{\text{circ}} \approx 10^4$ yr) due to viscous damping of tides induced in the convective outer envelope of the giant^{7,8}.

The rapid spin-up and persistent pulsed luminosity between bursts indicates that accretion is not inhibited by a centrifugal barrier (the propeller effect) at the disk–magnetosphere boundary. Standard accretion theory⁹ predicts that the magnetic field disrupts the disk and enforces co-rotation at a radius

$$r_m = \varepsilon (GM_x)^{-1/7} \mu^{4/7} \dot{M}^{-2/7} \quad (3)$$

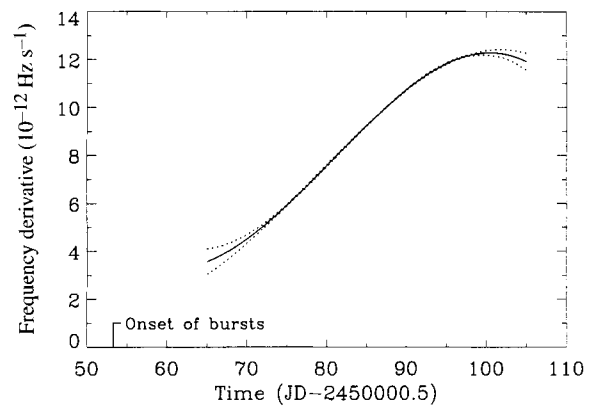


FIG. 3 The spin-up rate model determined by fitting the pulse phases. The dashed curves give 1σ errors for the value at any single point. The spin frequency of the neutron star derived from the fit is $\nu(t) = 2.141004031(14) + 9.228(27) \times 10^{-12} t + 1.88(58) \times 10^{-18} t^2 - 1.67(21) \times 10^{-25} t^3 + 1.12(30) \times 10^{-31} t^4$ Hz, where t is the emission time measured in seconds since JD 2450085.5.

where $\varepsilon \approx 0.5$ –1 (refs 10,11). The propeller effect occurs when the inner keplerian disk rotates more slowly than the magnetosphere. The requirement that the keplerian disk rotational frequency ν_K at r_m is greater than the neutron-star spin frequency, $\nu_K(r_m) > \nu$, results in a limit to the magnetic dipole moment,

$$\mu < \varepsilon^{-7/4} (2\pi\nu)^{-7/6} R^{1/2} L^{1/2} (GM_x)^{1/3} < 3 \times 10^{29} \text{ G cm}^3 \quad (4)$$

Here we assume $M_x = 1.4 M_\odot$, and a sub-Eddington luminosity L before the outburst. For a purely dipole field, and $R = 10^6$ cm, the field strength at the polar cap will be $B_d \lesssim 6 \times 10^{11}$ G, placing the cyclotron line energy below 7 keV. This modest field strength is consistent with arguments supporting accretion-induced magnetic field decay in neutron stars^{12–14}.

Two types of bursts have been found in X-ray bursters: type I due to nuclear burning of accreted material, and type II, thought to be due to an accretion instability. The type II bursts have been seen only in the Rapid Burster, MXB1730–335. For GRO J1744–28, the nuclear burning mechanism has been eliminated by energetics arguments². The bursts seen in GRO J1744–28 have some similarities with the type II bursts seen in the Rapid Burster^{2,15}, and are also likely to involve some unknown accretion instability.

All models we have examined for the Rapid Burster type II bursts, however, are inconsistent with our observations. The persistent X-ray pulsations rule out models that posit especially weak fields, $B \lesssim 10^8$ G (refs 16–18), and those that require the magnetic and spin axes to be aligned^{19,20}. Our limit (equation (4)) to the magnetic dipole field rules out models requiring ultra-strong dipole fields, $B_d \approx 10^{15}$ G (ref. 21). The star cannot be in the rapid propeller regime²², nor can the inner disk be rotating at nearly the same rate as the neutron star throughout the outburst as required by some models^{11,23,24}. Moreover, in the centrifugal gating mechanism of Spruit and Taam¹¹, the star should be spinning down between bursts when the centrifugal barrier at the magnetosphere reduces the accretion flow.

We have established GRO J1744–28 to be a neutron star accreting from an accretion disk fed by a low-mass companion. Our observations rule out many models for the burst mechanism, but the question of where in the accretion flow an instability occurs that results in the bursts remains unanswered. The answer could lie with the physical processes occurring in the boundary region between the disk and the magnetosphere. If so, then a successful model of GRO J1744–28 promises to provide new insights into this poorly understood region common to many neutron-star binary systems. $\square$

Received 8 February; accepted 10 April 1996.

1. Fishman, G. J. *et al.* *IAU Circ. No.* 6272 (1995).
2. Kouveliotou, C. *et al.* *Nature* **379**, 799–801 (1996).
3. Fishman, G. J. *et al.* in *Gamma Ray Observatory Science Workshop* (ed. Johnson, W. N.) 39–50 (NASA, Goddard Space Flight Center, Greenbelt, MD, 1989).
4. White, N. E., Swank, J. M. & Holt, S. S. *Astrophys. J.* **270**, 711–734 (1983).
5. Ftaclos, C., Kearney, M. W. & Pechenick, K. *Astrophys. J.* **300**, 203–208 (1986).
6. Paczynski, B. *Acta astr.* **17**, 287–296 (1967).
7. Phinney, E. S. & Kulkarni, S. R. *A. Rev. Astr. Astrophys.* **32**, 591–639 (1994).
8. Zahn, J. P. *Astr. Astrophys.* **57**, 383–394 (1977).
9. Henrichs, H. F. in *Accretion Driven Stellar X-ray Sources* (eds Lewin, W. H. G. & van den Heuvel, E. P. J.) 393–429 (Cambridge Univ. Press, 1983).
10. Ghosh, P. & Lamb, F. *Astrophys. J.* **232**, 259–276 (1979).
11. Spruit, H. C. & Taam, R. E. *Astrophys. J.* **402**, 593–604 (1993).
12. Blondin, J. M. & Freese, K. *Nature* **323**, 786–788 (1986).
13. Taam, R. E. & van den Heuvel, E. P. J. *Astrophys. J.* **305**, 235–235 (1986).
14. Romani, R. W. *Nature* **347**, 741–743 (1990).
15. Lewin, W. H. G., Rutledge, R. E., Kommers, J. M., Van Paradijs, J. & Kouveliotou, C. *Astrophys. J. Lett.* (in the press).
16. Milgrom, M. *Astr. Astrophys.* **172**, L1–L2 (1987).
17. Hanawa, T., Hirokuni, K. & Kawai, N. *Astrophys. J.* **336**, 920–923 (1989).
18. Walker, M. A. *Astrophys. J.* **385**, 651–660 (1992).
19. Hayakawa, S. *Phys. Rep.* **121**, 317–406 (1985).
20. Kuipers, J. & Kuipers, M. *Astr. Astrophys.* **286**, 491–496 (1994).
21. Hanami, H. *Mon. Not. R. astr. Soc.* **233**, 423–436 (1988).
22. Singh, L. M. & Duorah, H. L. *Astrophys. Space Sci.* **92**, 143–148 (1983).
23. Baan, W. A. *Astrophys. J.* **214**, 245–250 (1977).
24. Baan, W. A. *Astrophys. J.* **277**, 987–994 (1979).

ACKNOWLEDGEMENTS. We thank J. van Paradijs for useful discussions, D. Tahmouh for conducting searches in archival BATSE data for prior outburst of GRO J1744–28, and C. Wilson for analysing occultation data for pulse fraction determination.

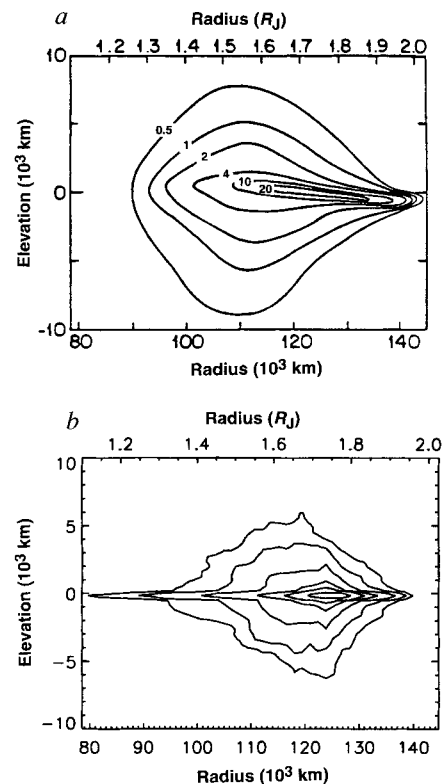


FIG. 1 *a*, The intensity contours of a halo slice reprojected into a rectangular coordinate frame in units of $\alpha\omega_0 P \times 10^9 \text{ km}^{-1}$, where α is the linear extinction coefficient, ω_0 is the single-scattering albedo and P is the phase function. The original Voyager 2 image (Flight Data System no. 20691.27) was taken through a violet filter (wavelength $\lambda = 0.431 \mu\text{m}$) at an average phase angle (Sun–ring/halo–camera) of $\theta = 175.8^\circ$. *b*, Calculated contour plots of the normalized brightness distribution ($B/\max(B)$) for dust particles with density $\rho = 3 \text{ g cm}^{-3}$ assuming an exponent $\gamma^* = -5.5$ in the size distribution of the dust production. The contour levels are 0.0125; 0.025; 0.05; 0.1; 0.25 and 0.5.

METHODS. To generate the results shown in *b*, we have numerically followed particles with radius $a = 0.1 \mu\text{m}$ (10,000 particles); 0.2 (10,000); 0.5 (1,000); 1 (1,000); 2 (100); 5 (100) and 10 (10), where in parentheses we note the number of calculated trajectories with randomly selected initial conditions. For each run we generated a random longitude and a position in the main ring to imitate the assumed constant surface density of the macroscopic source bodies. As the integration proceeded for each particle of a given grain size, we continuously updated a matrix, $N(i, j)$, whose elements record the number of hits a particle had in a box of size Δr and Δz at position $r = r_{\min} + i\Delta r$ and $z = z_{\min} + j\Delta z$. Our 50×200 matrix covered the range $1R_J < r < 2R_J$ and $-0.5R_J < z < 0.5R_J$ with a resolution of $\Delta r = 1R_J/50 \approx 1,400 \text{ km}$ and $\Delta z = 1R_J/200 \approx 350 \text{ km}$. For each particle size (index k), the matrices from individual trajectories were averaged: $N_k(r, z) = (\sum_{i=1}^{n_k} N(i, j))/n_k$, where n_k is the number of runs for the k th particle size. The N_k matrices were used to calculate the 'real' particle number density, $D_k(r, z)$, for each particle size: $D_k(r, z) = N_k P_k \Delta t / (2\pi r \Delta z \Delta r)$, where P_k is the production rate of the particles in the k th size bin, and Δt is our sampling rate (1 minute). The number densities in turn can be used to estimate the optical depth and brightness using a Mie-scattering code to calculate the optical properties of the grains (extinction and scattering coefficients, and the phase function) assuming an index of refraction $m = 1.5 - i0.01$ and the λ and θ as in *a*.

system is a much fainter outward extension of the main ring which almost reaches the orbit of the moon Amalthea⁵. Our understanding of the spatial structure of the ring/halo region is mainly based on the processing technique whereby images are transformed into intensity contours of cross-section slices⁴ (Fig. 1*a*).

Common to all theoretical models—including the one presented here—is the assumption that the large parent bodies in the main ring continuously supply the small dust particles via bom-

The structure and dynamics of Jupiter's ring

Mihály Horányi* & Thomas E. Cravens†

* Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309-0392, USA

† Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA

JUPITER'S ring has posed a problem since it was first discovered 15 years ago. Its inner edge flares into a torus, the shape of which has not yet been accurately modelled, although it was recognized early on^{1–3} that electromagnetic effects might be important in determining the spatial and size distribution of the dust particles comprising the ring. These early models suggested that sulphur and oxygen ions dominate the plasma environment of the ring; as these ions diffuse inwards from Io, they produce negatively charged dust grains in the ring. Here we propose that Jupiter's ionosphere is the dominant source of plasma and that the low plasma density allows ultraviolet radiation from the Sun to photoionize the grains, giving them a positive charge. The resulting gradient in the equilibrium charge distribution transports the grains rapidly—on surprisingly short timescales of hours to days—towards Jupiter. The brightness distribution resulting from this model matches closely the observed brightness distribution, suggesting that our model captures the most important processes that shape this ring.

We start with a summary of the observations⁴. The main ring is located at $1.71R_J \leq r \leq 1.81R_J$ (Jupiter's radius $R_J \approx 71,398 \text{ km}$) and has an optical depth of $\tau \approx 10^{-5}$. It consists of "large bodies" in the centimetre to metre (perhaps even kilometre) size range and a large population of dust particles in the $0.1 \leq a \leq 100 \mu\text{m}$ size range with approximately equal contributions to the optical depth and a scale height of $\sim 300 \text{ km}$. The inner edge of the main ring "expands" into a vertically extended, torus-shaped halo with a maximum scale height of 3,000 km. The halo gradually fades from the images near an inner radial distance of $\sim 1.4R_J$. The dust size distribution in the ring/halo follows a power law: $n(a)da \propto a^{-\gamma}da$, where a is the grain radius and $n(a)da$ is the number of grains in the size range a to $a + da$. The spectral index consistent with the observations⁴ is $\gamma = -2.5 \pm 0.5$. The third component of the ring

bardment by micrometeoroids and/or dust originating from Io. The ejected dust grains are exposed to the magnetospheric plasma as well as to sunlight, and consequently acquire electrical charge. The motion of charged grains is influenced by the Lorentz force due to Jupiter's magnetic field in addition to gravity, drag and radiation pressure¹⁻³.

In previous models the equilibrium surface potential of small dust grains was estimated to be about -10 volts, due to the assumption that the rings exist in a plasma environment produced by Io (+1 volt surface potential on a 1- μ m grain is a result of 700 missing electrons; for isolated grains, the potential ϕ and the charge Q are related as $\phi = Q/a$). Plasma drag, fluctuations of the magnetic field and the grain's electrostatic charge, and resonance forcing by the higher-order terms in the magnetic field, have all been invoked to explain dust transport from the main ring¹⁻³. None of these models yielded a quantitative match with the observations.

Our model involves an alternative plasma source: photoelectrons escaping from Jupiter's ionosphere⁶⁻⁸. Assuming that these electrons, with a characteristic energy $E \approx 10$ eV, are free to stream up the magnetic field lines, the net flux at the equatorial plane (that is, from both hemispheres) is $F \approx (6 \times 10^6) L^{-3} \text{ cm}^{-2} \text{ s}^{-1}$, where $L = r/R_J$ for the assumed dipole field. Ionospheric protons (or other ions) must accompany this electron flux so that quasi-neutrality can be maintained, thus creating an ambipolar electric field with an associated potential drop between the planet and the ring plane. This ambipolar field could reduce the electron flux by a factor of ~ 2 . In our calculations we also used photoelectron fluxes that are a factor of 10 lower and higher than our nominal flux to show that this uncertainty does not change our conclusions. The transient plasma clouds generated during micrometeoroid impacts⁹ represent a negligible plasma source owing to the low optical depth of Jupiter's ring.

Owing to the low densities, plasma drag is not important and photoelectron production from the dust, due to the solar ultra-

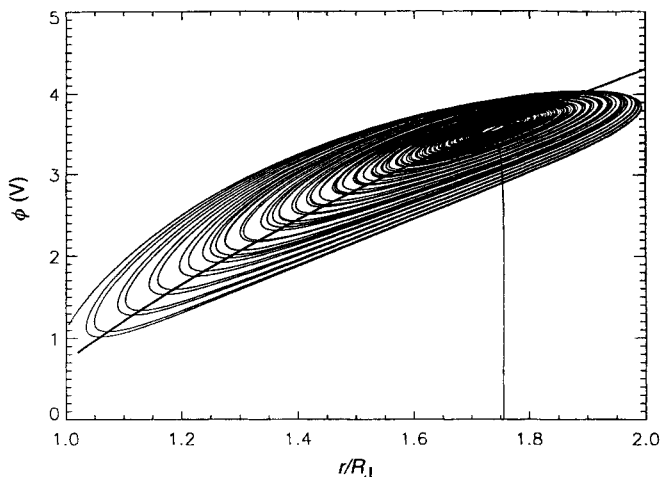


FIG. 2 The surface potential (ϕ) of an initially uncharged dust particle with radius $a = 1 \mu\text{m}$ and density $\rho = 3 \text{ g cm}^{-3}$ (thin line) and the equilibrium potential ϕ_{eq} (thick line), versus distance from Jupiter. The dust grain was started on a circular Kepler orbit at $r = 1.75R_J$. The grain charge is determined by the balance between the flux of ionospheric photoelectrons F to, and the flux of photoemitted electrons f from, the dust particles. The flux of photoelectrons emitted from an uncharged (or negatively charged) dust grain $f = 2.5 \times 10^{10} \kappa d^{-2} \text{ cm}^{-2} \text{ s}^{-1}$, where d is the distance from the Sun in AU and κ is a yield parameters¹⁸. For an insulating ring particle ($\kappa \approx 0.1$) orbiting Jupiter, $f \approx 10^8 \text{ cm}^{-2} \text{ s}^{-1}$. Because $f > F$, the grains charge positively, reducing f and enhancing F , until these fluxes balance as the grain's charge approaches its equilibrium value. Owing to the capacitance of the dust grains, their time-dependent charge differs from the equilibrium value.

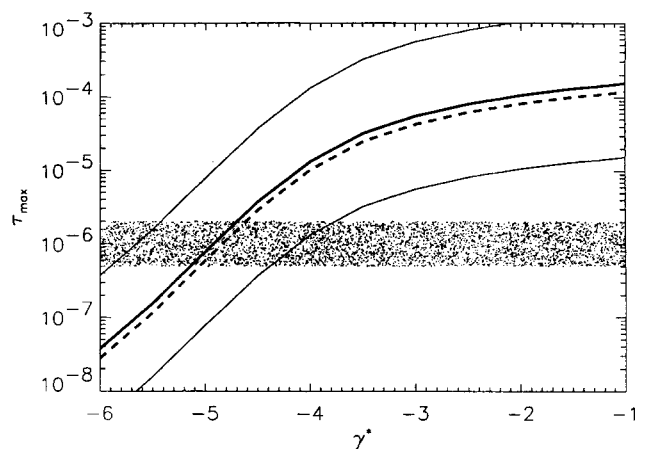


FIG. 3 The calculated maximum normal optical depth τ_{max} (at $r \approx 1.8R_J$) due to particles in the size range of $0.1 \leq a \leq 10 \mu\text{m}$, with a density of $\rho = 1$ (dashed line), 3 (continuous lines) as function of the exponent in the ejecta size distribution function, γ^* . The total mass production rate from the main jovian ring is $\dot{M} = F_M 2AY \tau_b \approx 300 \text{ g s}^{-1}$ (thick lines), where $F_M \approx 10^{-16} \text{ g cm}^{-2} \text{ s}^{-1}$ is the mass flux of the interplanetary meteorites^{19,20}, $A \approx 5 \times 10^{19} \text{ cm}^2$ is the geometrical surface area of the ring with a factor of 2 to account for both of its sides, $Y = 10^4$ is the expected ratio of the produced and impacting mass fluxes^{19,20}, and $\tau_b \approx 3 \times 10^{-6}$ is the optical depth of the macroscopic bodies in the main ring. Most of these parameters are uncertain and $\dot{M}$ could be at least an order of magnitude higher or lower. Accordingly we have also marked the expected higher and lower optical depth (thin lines) for $\dot{M} = 3 \text{ kg s}^{-1}$ and $\dot{M} = 30 \text{ g s}^{-1}$, respectively. With the largest $\dot{M}$ considered here for the entire main ring, source bodies with an initial radius of $\sim 10 \text{ km}$ would erode away during the age of the Solar System. In all cases, the size distribution of the ejected dust was assumed to follow a power law¹⁸ in the size range of $0.1 \leq a \leq 100 \mu\text{m}$ so that $\dot{M} = (4\pi\rho/3) \int n(a)a^3 da = c(4\pi\rho/3) \int a^{3+\gamma^*} da$, where c is a normalization constant.

violet radiation, becomes the dominant charging process. As the flux of photoelectrons emitted from an uncharged dust grain (or an initially negatively charged grain emerging from the short-lived impact plasma cloud⁹), $f \approx 10^8 \text{ cm}^{-2} \text{ s}^{-1}$, exceeds the ionospheric electron flux, grains will charge positively. The radial gradient of the equilibrium potential ϕ_{eq} is also positive ($\partial\phi_{\text{eq}}/\partial r > 0$) owing to the radial dependence of F (Fig. 2). This radial dependence of the equilibrium potential plays a crucial role in our model.

To follow a dust particle we simultaneously integrate the equations of motion (accounting for Jupiter's gravity with its higher moments J_2 and J_4 , radiation pressure and electromagnetic forces) and the equation governing the variation of its electrical charge. Similar methods were used to study the Ulysses dust streams and the fate of dust from comet Shoemaker-Levy 9 at Jupiter¹⁰⁻¹². Newly released dust particles in the main ring charge positively, and start oscillating towards and away from Jupiter owing to their slightly eccentric orbits. As $\partial\phi_{\text{eq}}/\partial r > 0$, grains moving towards (away from) Jupiter—against (parallel to) the outward pointing co-rotational electric field—tend to have larger (smaller) charge than the equilibrium charge at any r owing to the charging time delay of these particles (Fig. 2). The grains swiftly lose energy and angular momentum¹³⁻¹⁵ and, consequently, they follow orbits with decreasing semimajor axes a_k , and increasing eccentricities e , crashing into Jupiter's atmosphere as soon as their perijove distance $a_k(1 - e) = 1$. We found that the orbital evolution of dust particles is surprisingly fast and their lifetime is proportional to their mass. The orbital period averaged energy loss is proportional to the difference in the grain's charge, ΔQ , during the inbound and the outbound leg of its trajectory. The charge $Q \propto a$, but the charging time $t_q \propto a^{-1}$. And because $\Delta Q \propto Qt_q$, the energy loss rate is approximately independent of a , and the lifetime $T \propto E/E \propto a^2$. The lifetime,

$T \approx 100(a/1 \mu\text{m})^3$ days, is an approximate fit to all our numerical results.

Particles spiralling towards Jupiter will not stay confined to the equatorial plane. The "tilted" nature of the jovian magnetic field¹⁶ results in out-of-the-ring-plane forces and torques. Particles moving around the equatorial plane will 'see' a component of the magnetic field, $B_{\parallel}$, that is parallel to the equatorial plane periodically changing its orientation. Particles on a prograde (anticlockwise) trajectory experience a Lorentz force pushing them above (below) the ring plane as they cross the sectors where $B_{\parallel}$ points towards (away from) Jupiter. This effect will force particles onto orbits with increasing inclinations³. The orbital evolution of a dust particle strongly depends on its size. The smallest grains, $a < 0.1 \mu\text{m}$, will spiral along magnetic field lines arriving at the top of the jovian atmosphere at high latitudes at the magnetic footprint of the ring¹¹. Particles in the size range of $0.1 < a < 1 \mu\text{m}$ attain moderate inclinations and will hit Jupiter at mid-latitudes; the large grains $a > 1 \mu\text{m}$, spiral inwards remaining close to the ring plane, and arrive at Jupiter close to the equator. This size sorting might have important implications for the latitude-dependent haze and cloud formation at Jupiter by delivering condensation seeds and allowing surface chemistry to proceed.

Figure 3 shows the calculated values for the maximum normal optical depth at $r \approx 1.8R_J$ as a function of γ^* , the power exponent of the ejected dust size distribution, which is generally believed to follow a power law¹⁷. With the uncertainties in the total ejecta mass production rate, our optical depth estimates fall in an 'acceptable' range for $-6 \leq \gamma^* \leq -4$. However, the spatial structure of the halo (Fig. 1a) puts further constraints on γ^* , as γ^* 'controls' the relative contribution of the different-sized particles with vastly different photometric weights. In fact, γ^* 'controls' the apparent spatial distribution of the halo by its weighting of different dust sizes, with small dust grains being more widely distributed than large ones. Figure 1b shows our calculated 'brightness' contours for $\gamma^* = -5.5$. The match with Fig. 1a is rather good. Smaller negative values of γ^* rapidly reduce the vertical extent of the halo. Note that $\gamma^* = -5.5$, combined with our lifetime estimate $T \propto a^3$, reproduces the observed size distribution⁴, as it is proportional to the product of the lifetime and the production rate, $n(a)da \propto T(a)\dot{n}(a)da \propto a^{-2.5}da$.

Several observations support our idea that the ring/halo region consists of very short-lived positively charged grains. Very short lifetimes could explain the controversial reports on the sporadic azimuthal brightness asymmetry of the main ring⁴. At any moment the ring brightness reflects the stochastic nature of the dust production mechanism, because the very short lifetimes do not allow for azimuthal averaging due to shear that longer-lived particles would experience. If this is the case, Galileo might observe sudden brightenings in the ring during the course of its mission. Furthermore, if interstellar grains contribute significantly to the bombarding flux, there could be a 12-year seasonal period (Jupiter's orbital motion) and a longer 22-year period (interstellar grain access to the Solar System is modulated by solar cycle) in the average brightness variations of the jovian ring. The suggested positive charge of small dust grains might be responsible for the bright features in the main ring that appear to be dynamically connected to the largest bodies Adrastea and Metis⁴. Positively charged grains start oscillating inward from their source producing brighter regions at the turning points.

Galileo will not make *in situ* observations near the ring. However, radio occultation measurements of the ionosphere and ultraviolet stellar occultation measurements of the neutral thermosphere will allow for better estimates of the ionospheric photoelectron flux to the ring plane. High-resolution images will reveal the spatial and size distribution of the larger source bodies in the main ring. Hence there will be an excellent opportunity to improve our model, and to test it by comparing its predictions with the expected new detailed observations of the spatial and size distribution of small dust particles of the halo region. □

Received 19 January; accepted 2 April 1996.

1. Consolmagno, G. J. *Nature* **285**, 557–558 (1980).
2. Morfill, G. E., Grün, E. & Johnson, T. V. *Planet. Space Sci.* **28**, 1087–1100 (1980).
3. Burns, J. A., Schaffer, L. E., Greenberg, R. J. & Showalter, M. R. *Nature* **316**, 115–119 (1985).
4. Showalter, M. R., Burns, J. A., Cuzzi, J. N. & Pollack, J. B. *Icarus* **69**, 458–498 (1987).
5. Showalter, M. R., Burns, J. A., Cuzzi, J. N. & Pollack, J. B. *Nature* **316**, 526–528 (1985).
6. Luhmann, J. G. & Walker, R. J. *Icarus* **44**, 361–366 (1980).
7. Waite, J. H. Jr et al. *J. geophys. Res.* **88**, 6143–6163 (1983).
8. Nagy, A. F., Barakat, A. R. & Schunk, R. W. *J. geophys. Res.* **91**, 351–354 (1986).
9. Morfill, G. E. & Goertz, C. K. *Icarus* **55**, 111–123 (1980).
10. Horányi, M., Morfill, G. E. & Grün, E. *Nature* **363**, 144–146 (1993).
11. Horányi, M., Morfill, G. E. & Grün, E. *J. geophys. Res.* **98**, 21245–21251 (1993).
12. Horányi, M. *Geophys. Res. Lett.* **21**, 1039–1042 (1994).
13. Burns, J. A. & Schaffer, L. E. *Nature* **337**, 340–343 (1989).
14. Northrop, T. G., Mendis, D. A. & Schaffer, L. E. *Icarus* **79**, 101–115 (1989).
15. Havnes, O., Morfill, G. E. & Melandso, F. *Icarus* **98**, 141–150 (1992).
16. Connery, J. E. P. *J. geophys. Res.* **98**, 18659–18679 (1993).
17. Grün, E., Morfill, G. E., Schwem, G. & Johnson, T. V. *Icarus* **44**, 326–338 (1980).
18. Whipple, E. C. *Rep. Prog. Phys.* **44**, 1197–1250 (1981).
19. Ip, W.-H. *J. geophys. Res.* **88**, 819–822 (1983).
20. Cuzzi, J. N. & Durisen, R. H. *Icarus* **84**, 467–501 (1990).

ACKNOWLEDGEMENTS. We have benefited from numerous discussions with members of the Planetary Rings Group at LASP, especially J. Colwell and L. Esposito. This work was supported by NASA.

A high-silica zeolite with a 14-tetrahedral-atom pore opening

C. C. Freyhardt*, M. Tsapatsis†, R. F. Lobo‡, K. J. Balkus Jr§ & M. E. Davis*

* Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

† Department of Chemical Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA

‡ Center for Catalytic Science and Technology, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716, USA

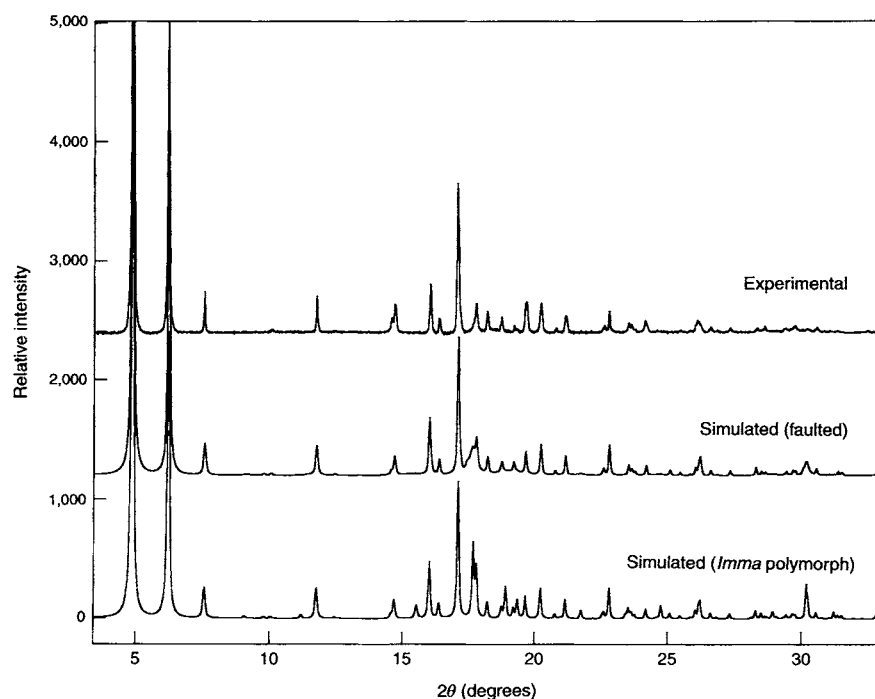
§ Department of Chemistry, University of Texas at Dallas, Richardson, Texas 75083, USA

ZEOLITES (microporous aluminosilicates) and related molecular sieves have found wide application as catalysts, sorbents and ion-exchange materials. New zeolites with large pores are much in demand^{1–4}, and have been sought for several decades^{4–7}. All known zeolites, both natural and synthetic, contain pores comprised of 12 or fewer tetrahedrally coordinated silicon or aluminium atoms (T-atoms), but several microporous aluminophosphates with wider pores are now known^{2,8–12}. The practical value of these large-pore phosphate-based materials is limited, however, by their poor thermal and hydrothermal stability. Here we report the synthesis of a high-silica zeolite with pores comprised of 14 T-atoms. Preliminary data indicate that this thermally stable large-pore material exhibits the kind of strong acidity that makes other zeolites useful catalysts.

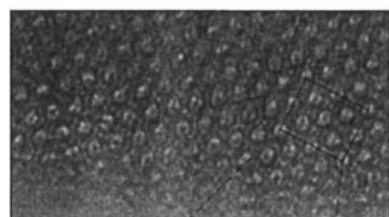
The first molecular sieve with pores containing rings of more than 12 T-atoms was the aluminophosphate VPI-5, synthesized in 1988⁸, which had 18-T-atom rings. Subsequently AlPO_4 ^{8,9,10}, cloverite², JDF-20¹¹ and ULM-5¹² were prepared, with pores of 14, 20, 20 and 16 T-atoms respectively. All of these materials are phosphate-based. In 1992 researchers at Mobil synthesized ordered silicate and aluminosilicate mesoporous materials¹³ with pores of 20–100 Å, up to ten times larger than those of existing molecular sieves. But because the inorganic portions of these mesoporous solids are not crystalline, their practical potential is limited¹⁴.

Our new material, which we call UTD-1 (University of Texas at Dallas number 1) was synthesized using bis-(pentamethyl-cyclopentadienyl)-cobalt(III) hydroxide, $[(\text{Cp}^*)_2\text{Co}]\text{OH}$, as a structure-directing agent¹⁵. The essentially pure-silica version of UTD-1 used here for the structure solution is prepared under typical

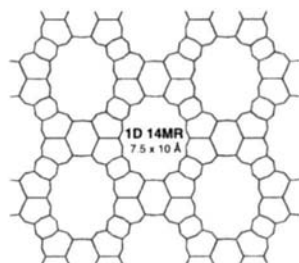
FIG. 1 Experimental (synchrotron radiation, $\lambda = 1.2513 \text{ \AA}$, top) and simulated (80% *Imma* polymorph + 20% *Bmmm* polymorph, middle; 100% *Imma* polymorph, bottom) powder X-ray diffractograms of organometallic-free UTD-1; the experimental diffractogram is shown after background subtraction.



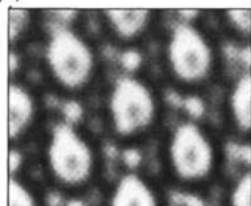
a Unprocessed HRTEM image



d Structure connectivity of UTD-1

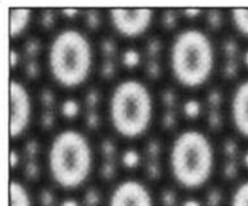


b



Processed image

c



Simulated image

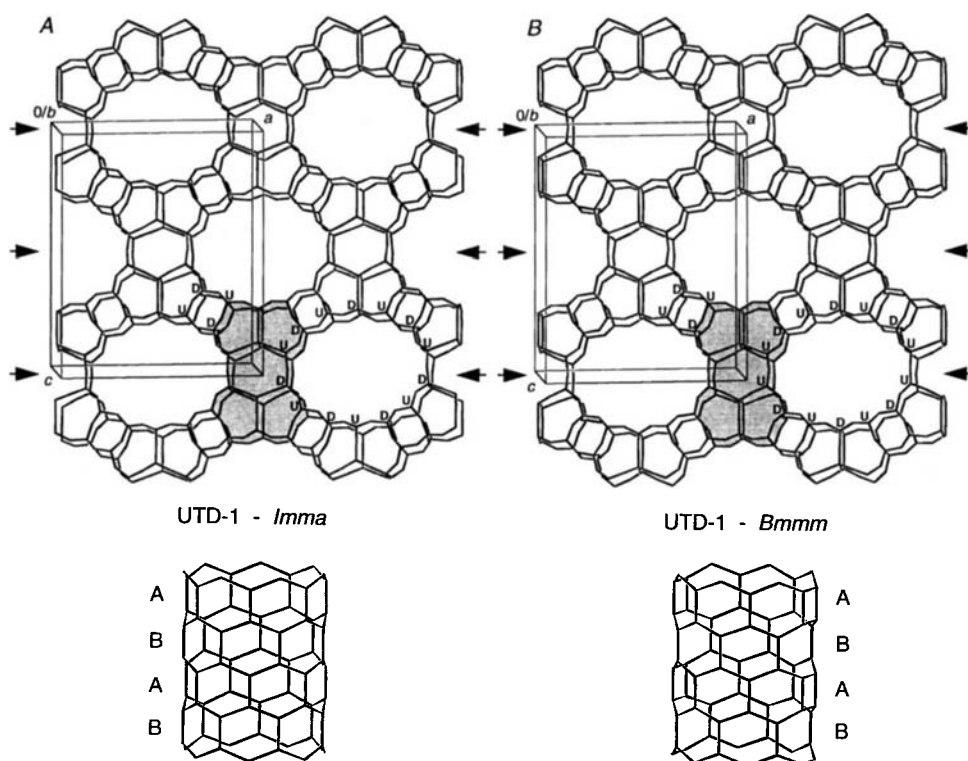
FIG. 2 Experimental (unprocessed, a; processed data, b) and simulated (c) high-resolution transmission electron micrographs (HRTEM) with topology of organometallic-free UTD-1 (d), as viewed along the large 14-MR channels ([010] zone axis). Samples for HRTEM were prepared by microtomy and images were recorded using a Philips 430 electron microscope operating at 200 kV.

hydrothermal conditions by heating a synthesis mixture of composition $0.13 [(\text{Cp}^*)_2\text{Co}]\text{OH} : 0.1 \text{ NaOH} : \text{SiO}_2 : 61 \text{ H}_2\text{O}$ to 175°C for 2 days in a sealed, Teflon-lined, stainless-steel autoclave. Upon calcination in air to 550°C , the organometallic complex $[(\text{Cp}^*)_2\text{Co}]^+$ is converted into cobalt oxide (grey coloured product). Removal of the cobalt oxide is accomplished by treatment of the sample with 12 N HCl for 2 hours at room temperature and afterwards with 7 N HCl at 50°C overnight. The white, solid product is subsequently washed with deionized water until the filtrate tests negative for chloride ions, and dried at 90°C for 4 hours.

The structure of UTD-1 was solved using an iterative process of model building and distance least-squares refinement (DLS-76¹⁶) of the atomic positions with subsequent comparison between the calculated (CERIUS¹⁷) and experimental synchrotron powder X-ray diffraction (SPXRD) pattern. A good match between the XRD data sets is obtained from a framework topology possessing

orthorhombic symmetry with the centred space group *Imma* (no. 74) (Fig. 1). However, there still remained discrepancies between the experimental and the calculated powder diffractograms. In particular, all *hkl* reflections with index $k = 2n + 1$ that appear in the simulated *Imma* pattern are missing in the experimental SPXRD diagram. Additionally, electron diffraction images of the organometallic-free UTD-1 recorded along the large pores containing 14-T-atom-membered rings (14-MRs) show alternating rows of sharp diffraction spots and streaked lines. These observations indicate some disorder and/or faulting of the framework along the direction of the 14-MR channels, that is, the [010] direction (see below), including a translation of $\pm 1/2$ *b*-axis. This kind of faulting does not cause a distortion of the straight 14-MR channels, as it is confirmed by the high-resolution transmission electron microscopy images, where no blocking of the large pores is observed (Fig. 2). The faulting occurs within the 002 planes and can be simulated by an intergrowth of two randomly distributed,

FIG. 3 Up(U)-down(D) connection sequence between structure layers and 14-MR channels in the pure *Imma* (A) and *Bmmm* (B) polymorph of organometallic-free UTD-1, which appear in the product at approximately 80% and 20%, respectively (one building unit is grey-shaded); arrows denote the position of the 002 faulting planes.



topologically closely related structure types (polymorphs). The simulations of the XRD patterns for the intergrowth structures were performed with the program DIFFaX¹⁸. For UTD-1, the two endmembers are related by consecutive shear shifts of $\pm 1/2 b$ along the 002 faulting planes that cause the framework structure to turn from an ordered *Imma* to an ordered *Bmmm* (no. 65, non-standard setting) polymorph or *vice versa* (Fig. 3). The best fit between the experimental and calculated XRD patterns for UTD-1 is achieved by using a faulting probability of 20%. Therefore, the material contains about 80% of the *Imma* and 20% of the *Bmmm* polymorph, respectively (see Fig. 1).

A unit-cell refinement based on the indexing of the reflections present in the experimental SPXRD pattern gives an orthorhombic unit cell with lattice parameters of $a = 18.981 \text{ \AA}$, $b = 8.415 \text{ \AA}$ and $c = 23.040 \text{ \AA}$ (unit-cell volume $V = 3,680.1 \text{ \AA}^3$). As there is no change in unit-cell size for the pure, ordered polymorphs *Imma* and *Bmmm*, the experimentally obtained cell parameters from the faulted UTD-1 are used for both polymorphs. In each case, the asymmetric unit contains 5 T-atoms and 12 oxygens, leading to a unit-cell content of $[\text{Si}_{64}\text{O}_{128}]$ (density $\rho = 1.73 \text{ g cm}^{-3}$) and a framework density of 17.4 T-atoms per $1,000 \text{ \AA}^3$. Table 1 contains the fractional atomic coordinates of the framework atoms for the *Imma* polymorph of UTD-1 free of structure-directing agent, as obtained from distance least-squares refinement (similar information for the *Bmmm* polymorph is available: see Supplementary Information). A detailed discussion of the faulting and a complete list of the structural data will be reported at a later date (C.C.F. *et al.*, manuscript in preparation).

Simulations of XRD patterns with other UD (up-down) sequences of T-atoms were also performed. Because of the faulting, unambiguous exclusion of alternative combinations of UD sequences is not possible at this time. A full discussion of UD sequences will be prevented elsewhere (C.C.F. *et al.*, manuscript in preparation).

Figure 2 illustrates the [010] projection of the UTD-1 structure along with the experimental and simulated high resolution transmission electron micrograph confirming the general topology of the structure. The structure of UTD-1 can be described as a fully condensed $\text{TO}_{4/2}$ network containing a one-dimensional pore

system running down the b -axis built of extra-large, elliptically-shaped pores circumscribed by 14-MRs. The main structural building unit within the pore walls consists of a double layer containing one 6-MR surrounded by four 5-MRs, as can be seen Fig. 3. Similar building units are found, for example, in the zeolites ZSM-48¹⁹, CIT-1²⁰ and others. Within each layer the building units in UTD-1 are linked via 4-MRs to form infinite, two-dimensional layers parallel to the a - c plane. Owing to the body-centring in the pure *Imma* polymorph, the layers are related by translational symmetry of $1/2 a$, $1/2 b$, $1/2 c$. For the ordered *Bmmm* polymorph, the symmetry relation between the layers contains only the translation components $1/2 a$ and $1/2 c$, as expected for a B-centred lattice. The faulting in the structure of UTD-1 can generally be described as occurrence of some disorder in the UD... connection sequence of the two-dimensional layers along the b -axis without a dislocation of the layers parallel to the a - c

TABLE 1 Fractional atomic coordinates for the *Imma* polymorph of organometallic-free UTD-1

Atom	Multiplicity	x	y	z
Si1	8	0.00000	0.56615	0.11668
Si2	8	0.00000	0.56613	0.25387
Si3	16	0.14910	0.43389	0.27618
Si4	16	0.22374	0.56621	0.16795
Si5	16	0.14353	0.43786	0.06494
O11	4	0.00000	0.75000	0.10080
O12	8	0.00000	0.54574	0.18528
O15	16	0.06843	0.48443	0.09034
O22	4	0.00000	0.75000	0.26973
O23	16	0.06859	0.48507	0.28021
O33	8	0.15603	0.25000	0.29102
O34	16	0.17785	0.46513	0.21236
O43	16	0.19478	0.53363	0.32116
O44	8	0.20777	0.75000	0.17707
O45	16	0.20365	0.51682	0.10341
O55	8	0.14989	0.50000	0.00000
O55a	8	0.15227	0.25000	0.06597

plane that would block the one-dimensional 14-MR pores. The main difference between the two polymorphs becomes apparent on the inner surface of the 14-MR pores as shown in Fig. 3. Whereas the pores of the *Imma* polymorph contain exclusively 6-MRs, the *Bmmn* polymorph pores possess 4-MRs and 8-MRs in addition to 6-MRs.

The approximate free diameter of the large 14-MR pore in UTD-1 is $7.5 \times 10 \text{ \AA}$ and leads to a geometrically calculated void volume (V_v) of $\sim 990 \text{ \AA}^3$ per unit cell (0.15 ml g^{-1}). This value is in reasonable agreement with the experimentally obtained volume of 0.13 ml g^{-1} ($V_v = 830 \text{ \AA}^3$ per unit cell) derived from N_2 adsorption data (isotherm available: see Supplementary Information). An estimate of the molecular volume (V_m) of the organometallic complex $[(\text{Cp}^*)_2\text{Co}]^+$ that is used as the structure-directing agent in the synthesis of UTD-1 gives a value of $V_m \approx 350 \text{ \AA}^3$. The comparison of this value with the available void volume in the structure implies that there are approximately two Co-complexes per unit cell in the as-synthesized product of UTD-1. This result agrees with the thermogravimetric analysis of the as-synthesized material that shows a weight loss of $\sim 13 \text{ wt\%}$ that can be

attributed to the decomposition of two $[(\text{Cp}^*)_2\text{Co}]^+$ complexes. One reason for the differences between the available pore volume and the molecular volume of the structure-directing agent complex is that the shape of the pore and the complex do not perfectly match.

Despite the existence of the extra-large, 14-MR-containing pores in the structure of UTD-1, the thermal stability of the framework of the organometallic-free material is remarkably high. Temperature-dependent powder XRD investigations show no recognizable loss in crystallinity below $1,000^\circ\text{C}$, even in a nitrogen atmosphere containing approximately 23 torr H_2O . Additionally, the aluminium-containing form of UTD-1 possesses Brønsted-acid sites of strength comparable to other high-silica zeolites, and is able to perform a broad range of hydrocarbon reactions²¹. Further details of acidity and catalytic behaviour of UTD-1 will be reported at a later date (C.C.F. *et al.*, manuscript in preparation). The range of interesting reactions catalysed by UTD-1 may also be expanded by the incorporation of other elements into the framework such as Ti^{22} , B, V and Zn (K.J.B. *et al.*, manuscript in preparation). □

Received 5 March; accepted 17 April 1996.

1. Davis, M. E. *Chem. Ind.* **4**, 137–139 (1992).
2. Estermann, M., McCusker, L. B., Baerlocher, C., Merrouche, A. & Kessler, H. *Nature* **352**, 320–323 (1991).
3. Davis, M. E. *Nature* **352**, 281–282 (1991).
4. Davis, M. E. *Nature* **337**, 117 (1989).
5. Barrer, R. M. & Villiger, H. Z. *Kristallogr.* **128**, 352–370 (1963).
6. Smith, J. V. & Dyrtch, W. J. *Nature* **309**, 607–608 (1984).
7. Brunner, G. O. & Meier, W. M. *Nature* **337**, 147–147 (1989).
8. Davis, M. E., Sakdarni, C., Montes, C., Garces, J. & Crowder, C. *Nature* **331**, 698–699 (1988).
9. Dessau, R. M., Schlenker, J. L. & Higgins, J. B. *Zeolites* **10**, 522–524 (1990).
10. Richardson, J. W. Jr & Vogt, E. T. C. J. *Solid St. Chem.* **87**, 469–471 (1990).
11. Jones, R. H. *et al.* *J. Solid St. Chem.* **102**, 204–208 (1993).
12. Loiseau, T. & Ferrey, G. J. *Solid St. Chem.* **111**, 403–415 (1994).
13. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
14. Davis, M. E. *Nature* **364**, 391–393 (1993).
15. Balkus, Jr, K. J., Gabrielov, A. G. & Sandler, N. *Mater. Res. Soc. Symp. Proc.* **368**, 369–375 (1995).
16. Baerlocher, C., Hepp, A. & Meier, W. M. *DLS-76: A Fortran Program for the Simulation of Crystal Structures by Geometric Refinement* (Institut fuer Kristallographie, ETH, Zürich, 1977).

17. *CERIUS Version 3.2* (Molecular Simulations, Cambridge, UK, 1993).
18. Treacy, M. M. J., Deem, M. W. & Newsam, J. M. *Proc. R. Soc. London.* **A 433**, 499–520 (1991).
19. Schlenker, J. L., Rohrbaugh, W. J., Chu, P., Valyocik, E. W. & Kokotailo, G. T. *Zeolites* **5**, 355–358 (1985).
20. Lobo, R. F. & Davis, M. E. *J. Am. chem. Soc.* **117**, 3766–3779 (1995).
21. Balkus, K. J. Jr, Gabrielov, A. G., Zones, S. I. & Chan, I. Y. in *Synthesis of Microporous Materials: Zeolites, Clays, Nanocomposites* (eds Kessler, H. & Occelli, M.) (Marcel Dekker, New York, in the press).
22. Balkus, K. J. Jr, Gabrielov, A. G. & Zones, S. I. *Stud. Surf. Sci. Catal.* **97**, 519–525 (1995).

ACKNOWLEDGEMENTS. We thank T. Vogt and D. E. Cox for their assistance in collecting the synchrotron powder XRD data, and the W. M. Keck Polymer Morphology Laboratory at the University of Massachusetts for the use of their electron microscopes. The data were collected at X7A beam line at the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY), that is supported by the Department of Energy, Division of Material Sciences and Division of Chemical Sciences. M.T. acknowledges support from the National Center for Electron Microscopy at Lawrence Berkeley Laboratory (NCEM/DOE fellowship). This work was supported by the Chevron Research and Technology Company (Richmond, California).

SUPPLEMENTARY INFORMATION. Available on Nature's World-Wide Web site <http://www.nature.com>. Paper copies are available from Mary Sheehan at the London editorial office of Nature.

Promoter-induced enhancement of the crystallization rate of zeolites and related molecular sieves

Rajiv Kumar, Asim Bhaumik, Ranjeet Kaur Ahedi & Subramanian Ganapathy

Catalysis and Physical Chemistry Division, National Chemical Laboratory, Pune-411008, India

ZEOLITES and other microporous molecular sieves^{1–4} exhibit useful catalytic, absorption and ion-exchange properties. Most synthetic zeolites are made by hydrothermal methods, which generally suffer from the drawback of long crystallization times⁵. Here we show that the crystallization rate of zeolites and related molecular sieves can be accelerated by a factor of as much as about five by the addition of various 'promoter' ions, such as perchlorate, phosphate, nitrate, sulphate and carbonate. By using enhanced-sensitivity ²⁹Si NMR studies to monitor soluble silicate species during crystallization, we show that the promoter ions enhance the rate of formation of the oligomeric (mainly tetrameric) silicate ions responsible for nucleation and growth of the crystals.

The addition of a small amount of the promoters ClO_4^- , PO_4^{3-} , NO_3^- , SO_4^{2-} , CO_3^{2-} , AsO_4^{3-} , ClO_3^- , BrO_3^- and IO_3^- significantly

enhances the crystallization process of a variety of zeolite structures (Fig. 1). These oxyanions can be used either as the acids or as the sodium or potassium salts, depending on the need to keep the molar gel composition constant. In a typical synthesis, a silica source such as fumed silica was first stirred with the required amount of template species (such as quaternary ammonium ions, which act as templates for pore formation) and alkali in deionized water for 1 h. Then a solution of the aluminium source (NaAlO_2) (43% Al_2O_3 , 39% Na_2O) in most of the cases, except in the case of zeolites FER, MTW and EU-1 where the source was aluminium sulphate) was added to the silica with vigorous stirring. Finally, an aqueous solution of the promoter in the form of an oxyacid or a salt was added to the gel and the stirring was continued for another 1 h at room temperature. Figure 1 indicates that, under otherwise similar synthesis conditions, the presence of promoters enhances nucleation and crystallization in the order: $\text{ClO}_4^- > \text{PO}_4^{3-} > \text{AsO}_4^{3-} \gg \text{none}$. Similar results were obtained in the case of other zeolites and various promoters (Table 1).

As the amount of promoter increased in the gel, the crystallization time decreased linearly initially and then remained the same after an optimum concentration. When plotted against the inverse of crystallisation time, the apparent energies of nucleation E_n and of crystallization E_c for ZSM-5, using ClO_4^- , PO_4^{3-} , AsO_4^{3-} and no promoter, increase in the order: $\text{HClO}_4 < \text{NaH}_2\text{PO}_4 < \text{Na}_2\text{HAsO}_4 \ll \text{none}$; the overall synthesis time is reduced in the reverse order. E_n and E_c were calculated according to the equation $\ln(1/\theta) = -E_n/RT$ and $\ln(1/k) = -E_c/RT$, where θ is the induction period and k is the half-crystallization time (that is, the time required to obtain 50% XRD crystallinity).

Scanning electron micrographs of various zeolite samples (not

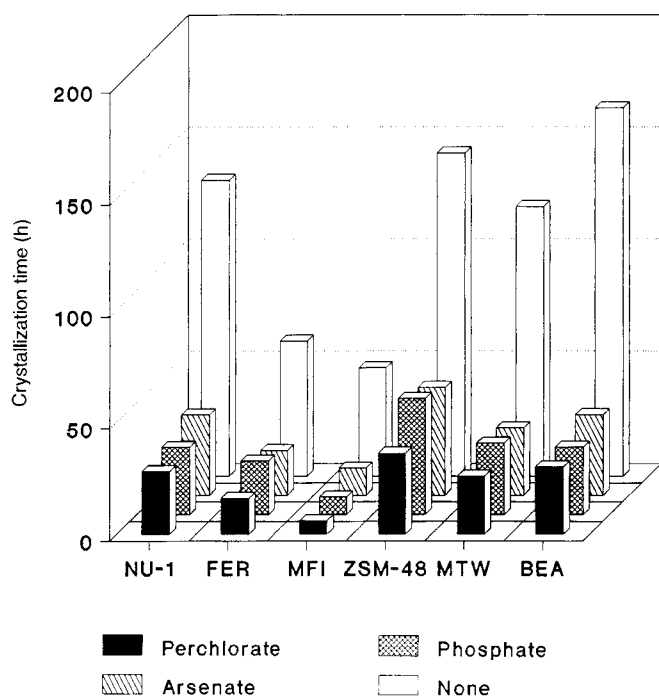


FIG. 1 Crystallization time (hours) of fully crystalline samples of various zeolites with and without using promoter P. Silica source: fumed SiO_2 , except for FER where sodium silicate was used. Molar gel compositions: NU-1, 40SiO_2 : Al_2O_3 : $5\text{Na}_2\text{O}$: 20TMAOH : $1,000\text{H}_2\text{O}$: 4P ($\text{pH} = 12.2 \pm 0.2$); FER, 80SiO_2 : Al_2O_3 : $24\text{Na}_2\text{O}$: 10pyrrolidine : $3,200\text{H}_2\text{O}$: 8P ($\text{pH} = 12.0 \pm 0.2$); MFI, 40SiO_2 : Al_2O_3 : $5\text{Na}_2\text{O}$: 10TPAOH : $1,200\text{H}_2\text{O}$: 4P ($\text{pH} = 10.8 \pm 0.2$); ZSM-48, 90SiO_2 : Al_2O_3 : $9\text{Na}_2\text{O}$: 15DIQ-6 : $2,700\text{H}_2\text{O}$: 9P ($\text{pH} = 11.6 \pm 0.2$); MTW, 120SiO_2 : Al_2O_3 : $10\text{Na}_2\text{O}$: 12DIQ-6 : $3,600\text{H}_2\text{O}$: 12P ($\text{pH} = 10.8 \pm 0.2$); BEA, 60SiO_2 : Al_2O_3 : $2\text{Na}_2\text{O}$: $0.5\text{K}_2\text{O}$: 30TEAOH : $600\text{H}_2\text{O}$: 6P ($\text{pH} = 12.5 \pm 0.2$); where TEA is tetraethyl ammonium, TPA is tetrapropyl ammonium, TMA is tetramethyl ammonium, DIQ-6 is hexamethylene bis (benzyl dimethyl ammonium hydroxide), P is HClO_4 , NaH_2PO_4 or Na_2HAsO_4 . The crystallization temperature was 170°C (NU-1, ZSM-48), 160°C (FER, MFI, MTW) and 140°C (BEA).

shown) demonstrated that the crystallite size of these samples is smaller and more uniform when the synthesis employs a promoter, probably due to the enhanced nucleation and crystallization rates. The ion-exchange capacity (K/Al molar ratio) of the ZSM-5 samples synthesized with and without NaH_2PO_4 was 0.93 and 0.91, respectively. Similarly, their *n*-hexane cracking activity (moles of *n*-hexane converted per mole of Al per h) was 18.0 and 17.2 respectively at 623 K, confirming that the material synthesized with promoters is comparable, if not better, in both textural and catalytic properties to those obtained by synthesis without promoters.

Our method is also applicable to Al-free pure silicalite-type and other metallo-silicates like titanium silicates (TS-1) and iron silicates. In the case of zeolites or other metallo-silicates, the promoter was not found in the solid (by chemical analysis) after repeated washing. However, in the case of Silicalite-1, an Al-free all-silica MFI-type analogue⁶, a small amount of one of the promoters, As(v), was found in the solid when the synthesis was carried out at lower temperature⁷ ($\sim 80^\circ\text{C}$). The TS-1 samples synthesised with and without promoter (H_3PO_4) were characterised by X-ray diffraction (XRD), ultraviolet-visible spectroscopy (strong charge-transfer band at 210 nm and no hump or broad peak in the range 280–380 nm) and infrared spectroscopy (960 cm^{-1} band along with other framework vibrations).

Liquid-state ^{29}Si NMR experiments on enhanced nucleation and crystallization of Si-ZSM-5 (Silicalite-1) using NaH_2PO_4 as promoter have allowed us to monitor the soluble precursors (Q^0 – Q^4 , mono-, di-, tri and tetra-meric silicate species, respectively)⁸ as a function of crystallization time spent during the hydrothermal

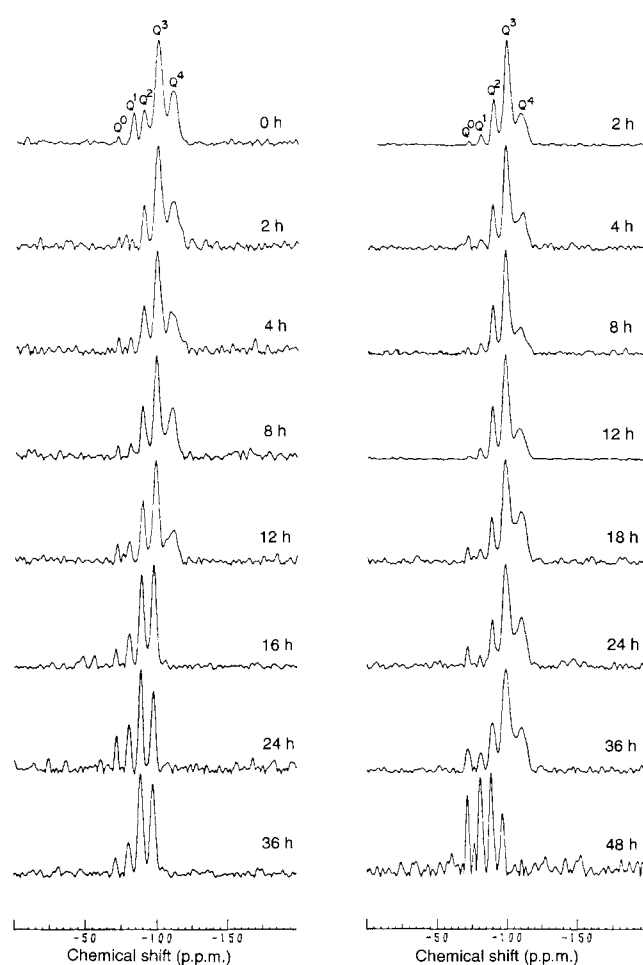


FIG. 2 ^{29}Si NMR spectra of Si-ZSM-5 reaction mixture with (left) and without (right) NaH_2PO_4 as promoter (reaction hours (at 85°C , samples were quenched and used immediately) are given next to each spectrum). Spectra were obtained at 59.601 MHz on a Bruker MSL-300 spectrometer, using a TMS standard. Co-addition of eight successive spin-echoes, generated by a CPMG pulse sequence^{11,12}, and a magnitude mode Fourier transform were used. Number of transients, 5,120; $\pi/2$ pulse width, $10.5\text{ }\mu\text{s}$; pulse sequence recycle time, 20 s; total accumulation time, 3.5 h. A home-made solenoidal coil assembly (silver-plated radio frequency coil having Q of 120) was used and the sample was placed in a sealed plastic tube (8 mm inside diameter, length 40 mm) devoid of any background ^{29}Si signal.

synthesis. Typical ^{29}Si NMR spectra of Si-ZSM-5 at varying degrees of crystallization in the presence of promoter (NaH_2PO_4) are shown in Fig. 2. The relative concentration of Q^4 silicate species (mainly double 3-, 4- and 5- rings of silicate tetrahedra⁸) (chemical shift $\delta = -110 \pm 1\text{ p.p.m.}$), among total soluble silicate species (Q^0 – Q^4), first increases slightly and then sharply decreases with increasing time, clearly suggesting that mainly Q^4 species are responsible for the condensation of silicate species to form crystalline solid. There is a striking correspondence between the increasing crystallinity of Si-ZSM-5, as shown by XRD, and the decreasing concentration of soluble Q^4 species measured by ^{29}Si NMR, the effect being dramatically fast in the presence of promoter anions (Fig. 3). These results demonstrate that the presence of promoter phosphate ions significantly accelerates both the formation of soluble Q^4 species and their rapid condensation into crystalline silicate units, which are not detected by liquid-state NMR under static conditions. Independent liquid-state ^{31}P NMR experiments (spectra not shown) conducted on Si-ZSM-5 in the presence of promoter (PO_4^{3-}) exhibit significant changes in ^{31}P signal position and resolution, and an analysis of

TABLE 1 Effects of promoters in zeolite synthesis

Zeolite	Temp. (°C)	Silica source	Crystallization time (h)		Template
			With promoter	No promoter	
ZSM-22	170	Fumed SiO ₂ *	24 (HClO ₄)	80	<i>N</i> -ethyl pyridinium
EU-1	160	Fumed SiO ₂	40 (HClO ₄)	144	Hexamethonium dibromide
Si-NCL-1	170	Fumed SiO ₂	18 (Na ₂ HAsO ₄)	56	Hexamethylene bis (triethyl ammonium)
BEA	140	Fumed SiO ₂	40 (NaNO ₃) 44 (NaHCO ₃) 48 (Na ₂ SO ₄) 60 (KBrO ₃) 96 (KIO ₃)	156	Tetraethyl ammonium hydroxide
Y	100	Sodium silicate†	4 (Na ₂ HPO ₄)	11	None
TS-1	170	TEOS‡	4 (H ₃ PO ₄)	24	Tetrapropyl ammonium hydroxide
	160		6 (H ₃ PO ₄)	30	
	140		10 (H ₃ PO ₄)	—	
	120		24 (H ₃ PO ₄)	—	

* Sigma S-5130, USA.

† 28% SiO₂, 8.5% Na₂O.

‡ Tetraethyl orthosilicate, Aldrich, USA.

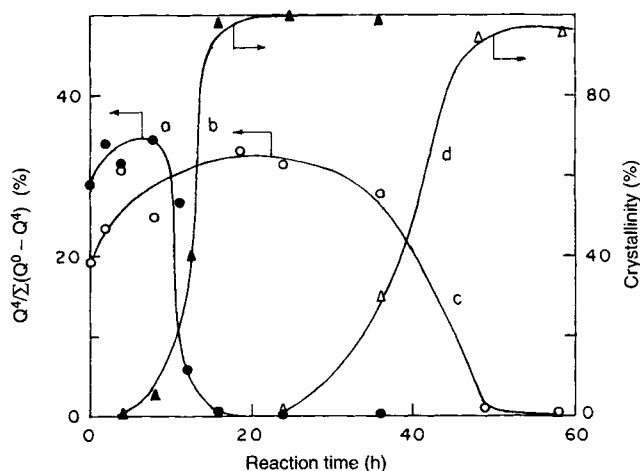


FIG. 3 Reaction-time dependence of relative concentration of Q⁴ species (from ²⁹Si NMR; left-hand ordinate) and percentage crystallinity of the same samples (from XRD; right-hand ordinate). Curves a (filled circles) and b (empty circles) for samples with promoter; curves c (filled triangles) and d (empty triangles) for samples without promoter.

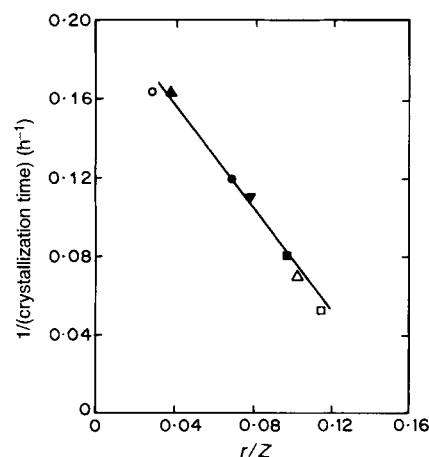


FIG. 4 Plot of crystallization time against charge/radius ratio of central cation in promoter NO₃⁻ (○), ClO₄⁻ (▲), PO₄³⁻ (●), ClO₃⁻ (▼), AsO₄³⁻ (■), BrO₃⁻ (△) and IO₃⁻ (□) in the synthesis of zeolite ZSM-5.

these, in turn, suggests a catalytic role played by the promoter in enhancing zeolite crystallization.

Knight *et al.*⁹ and Davis¹⁰ have argued in favour of the clathration hypothesis of zeolite formation, whereby the silicate polyanions (mainly Q⁴ species) clathrate the organic structure-directing agent (TPA, tetrapropyl ammonium, in Si-ZSM-5 system). These clathrated TPA-silicate hydrophobic hydration spheres come closer and overlap, thus forming composite species leading to crystal growth¹⁰. It seems that the presence of promoter oxyanions greatly polarizes the hydrophobic hydration spheres formed around silicate units and the alkyl groups of the organic structure-directing agent, thereby facilitating the overlapping of TPA-enclathrated silicate polyanions. Such overlapping, in turn, forms the composite species at the onset of crystallisation. In support of this view, we observed a good correlation between the *Z/r* (charge/radius ratio) of the central cation of promoter oxyanions and the corresponding crystallization time of ZSM-5 synthesized in the presence of such promoters (Fig. 4). As the polarizing ability (greater *Z/r* value) of the oxyanions increases, the crystallization becomes faster. These results suggest that the

presence of ionic species with greater polarizing ability enhances the process of crystallization dramatically by speeding up the condensation process. □

Received 1 August 1995; accepted 15 April 1996.

1. Barrer, R. M. *Hydrothermal Chemistry of Zeolites* (Academic, New York, 1982).
2. Breck, D. W. *Zeolite Molecular Sieves* (Wiley, New York, 1974).
3. Szostak, R. *Molecular Sieves: Principles of Synthesis and Identification* (Van Nostrand Reinhold, New York, 1989).
4. Karge, H. G. & Weitkamp, J. (eds) *Zeolites as Catalysts, Sorbents and Detergent Builders: Applications and Innovations* (Elsevier, Amsterdam, 1989).
5. Inui, T. in *Zeolite Synthesis* Ch. 33 (eds Occelli, M. L. & Robson, H. E.) 479–492 (ACS Symp. Ser. 398, Am. Chem. Soc., Washington DC, 1989).
6. Flanigen, E. M. *et al.* *Nature* **271**, 512–516 (1978).
7. Bhaumik, A., Hegde, S. G. & Kumar, R. *Catal. Lett.* **35**, 327–334 (1995).
8. Engelhardt, G. & Michel, D. *High-Resolution Solid-State NMR of Silicates and Zeolites* Ch. 3, 75–105 (Wiley, New York, 1987).
9. Knight, C. T. G., Syvitski, R. T. & Kinrade, S. D. *Stud. Surf. Sci. Catal.* **97**, 483–488 (1995).
10. Davis, M. E. *Stud. Surf. Sci. Catal.* **97**, 35–44 (1995).
11. Carr, H. Y. & Purcell, E. M. *Phys. Rev.* **94**, 630–638 (1954).
12. Meiboom, S. & Gill, D. *Rev. Sci. Instrum.* **29**, 688–691 (1958).

ACKNOWLEDGEMENTS. We thank A. A. Belhekar for SEM/EDX analysis of the samples. A.B. thanks CSIR, New Delhi for a Senior Research Fellowship.

A welfare-based index for assessing environmental effects of greenhouse-gas emissions

James K. Hammitt*, Atul K. Jain†, John L. Adams‡
& Donald J. Wuebbles†

* Center for Risk Analysis and Department of Health Policy and Management, Harvard School of Public Health, 718 Huntington Avenue, Boston, Massachusetts 02115, USA

† Department of Atmospheric Sciences, University of Illinois, 105 South Gregory Avenue, Urbana, Illinois 61801, USA

‡ RAND Corporation, 1700 Main Street, Santa Monica, California 90407 USA

THE global warming potential (GWP) index¹⁻³ is a measure of the relative contribution of the emissions of different greenhouse gases to the radiative forcing of the atmosphere—and thus to climate change—over a given time period. But this index does not represent the effects of climate change, and therefore does not provide an adequate basis for policy decisions about emissions reductions. This inadequacy has led to the proposal of an alternative, the economic-damage index (EDI)⁴⁻⁷. This index compares the effect of different greenhouse-gas emissions on global economic welfare. Here we use a simple climate model to calculate the EDIs for a range of climate-change/greenhouse-gas emission scenarios, and compare the values to the corresponding GWPs. Although the values of these indices are, at this stage, broadly similar in both magnitude and uncertainty, the prospects of reducing these uncertainties by future research are better for the EDI.

An index of the relative effects on economic welfare of unit mass emissions of alternative greenhouse gases (GHGs) is useful for evaluating trade-offs between GHG emissions for policy making, international negotiation, and for determining the most cost-effective means for limiting contributions to climate change. Such an index can also be used in a decentralized regulatory mechanism that limits the total effect of a nation's or firm's emissions, but permits the nation or firm to allocate its emissions among GHGs and to trade emission credits with others. (In an analogous fashion, the Montreal Protocol permits nations to reallocate their consumption of ozone-depleting substances, within classes, subject to a limit on total consumption weighted by ozone depletion potential). Because decentralized mechanisms do not require the regulator to know national or firm-specific costs of emission reductions, such mechanisms can substantially reduce abatement costs⁸ if difficulties in measuring GHG sources and sinks can be overcome⁹.

The GWP for GHG_{*i*} is defined as the ratio of the partial derivative of the time-integrated radiative forcing with respect to GHG_{*i*} relative to the partial derivative with respect to a standard GHG (by convention, CO₂)¹⁻³

$$\text{GWP}_i = \frac{\frac{\partial}{\partial e_i} \left[\int_0^h F(t) dt \right]}{\frac{\partial}{\partial e_0} \left[\int_0^h F(t) dt \right]} = \frac{\int_0^h \Delta C_i(t) R_i(t) dt}{\int_0^h \Delta C_0(t) R_0(t) dt} \quad (1)$$

where e_i and e_0 denote unit mass emissions of GHG_{*i*} and of CO₂ at time 0, h is a specified time horizon, and $F(t)$ denotes total radiative forcing at time t , respectively. In the second formulation, $\Delta C_i(t)$ and $\Delta C_0(t)$ are the differences in atmospheric concentration of GHG_{*i*} and of CO₂ between scenarios with and without incremental emissions of each GHG at time 0, and $R_i(t)$ and $R_0(t)$ are the incremental radiative forcing per unit of each GHG (the dependence of radiative forcing on atmospheric composition, and

hence t , can be neglected when GWP is evaluated for a constant-composition atmosphere¹⁻³). The GWP calculation treats GHGs as if their effects depend only on cumulative radiative forcing before the time horizon h . In an alternative formulation, the contribution to the index of future incremental forcing is reduced by exponential discounting and the horizon is extended to infinity¹⁰.

From a policy or economic-welfare perspective, radiative forcing *per se* is not important; what is significant is the effect of the associated climate change on human and natural systems that affect economic welfare (the 'damages', positive or negative, associated with climate change). To illustrate this, we consider a simple example in which climate-induced damages depend only on the maximum global annual-mean surface temperature reached¹¹. An incremental emission of a gas with a short atmospheric-residence time like CH₄ has negligible effect compared with an emission of a long-lived GHG like CO₂ if the emissions occur far enough before the year in which mean temperature peaks. As the emission date approaches the year of the peak, the emission of CH₄ becomes more important relative to that of CO₂. Neither GWP nor exponentially discounted radiative forcing¹⁰ properly account for the relative importance of short- and long-lived GHGs in this example, or in more realistic settings.

The EDI takes account of relative increments in economic welfare. In addition to climate-induced effects, EDI can incorporate effects that occur through other pathways including stratospheric-ozone depletion by halocarbons and enhanced plant growth by CO₂ fertilization⁵. EDI depends on future GHG emissions, because atmospheric composition affects incremental radiative forcing, some atmospheric-removal processes, and the economic effects of incremental climate change. (GWP also depends on future GHG emissions¹², although it is conventionally evaluated for a constant atmosphere.) Unlike GWP, EDI also depends on other factors that affect climate change (for example, H₂O feedback and sulphate aerosols) and factors that influence climate-induced effects and their value to humans. Along a socially optimal GHG-emission trajectory, the values of EDI correspond to ratios of the optimal marginal social costs or shadow prices of GHG emissions^{4,5}.

The EDI for GHG_{*i*} is defined as the reduction in emission of a standard GHG (CO₂) required to offset the incremental damage that would otherwise accompany an increased emission of GHG_{*i*}. Equivalently, EDI_{*i*} is the partial derivative of the present value of economic welfare loss with respect to emissions of GHG_{*i*} relative to the partial derivative with respect to CO₂ emissions

$$\text{EDI}_i = \frac{\frac{\partial}{\partial e_i} W[C(t)]}{\frac{\partial}{\partial e_0} W[C(t)]} = \frac{\int_0^\infty \Delta C_i(t) \lambda_i(t) dt}{\int_0^\infty \Delta C_0(t) \lambda_0(t) dt} \quad (2)$$

where $W[C(t)]$ represents economic welfare loss due to the time path of GHG concentrations $C(t)$. In the second formulation, $\lambda_i(t)$ denotes the marginal social cost or shadow price of an additional unit concentration of GHG_{*i*}.

For simplicity and consistency with current integrated-assessment models¹³⁻¹⁵, we neglect non-climate effects of GHG emissions and assume that damages due to climate change (net of adaptation costs) can be represented by

$$W[\Delta T(t)] = \int_0^\infty \left(\frac{1}{1+r} \right)^t \alpha \text{GDP}(t) D[\Delta T(t)] dt \quad (3)$$

where r is the discount rate, α is a scaling constant, GDP(t) is gross world product and $\Delta T(t)$ is the increase in global annual-mean surface temperature from its value in 1990. The level and shape of the damage function are highly uncertain. The constant α cancels in EDI, so uncertainty about its value (assuming it to be positive) is unimportant. (If non-climate impacts such as CO₂

TABLE 1 EDI for selected parameter values, atmospheric lifetime and GWP

EDI		CO ₂	N ₂ O	CH ₄	CFC-11	CFC-12	HCFC-22
Middle case*		1	354.8	11.0	3,123	9,067	773
Discount rate, <i>r</i>	1% yr ⁻¹	1	322.2	3.73	1,833	7,950	261
	5% yr ⁻¹	1	366.2	23.70	4,317	9,596	1,688
Damage-function exponent, γ	1	1	354.7	27.21	4,182	9,279	1,962
	3	1	340.1	5.10	2,334	8,527	353
	'Hockey-stick'†	1	319.4	6.07	2,043	7,910	428
Climate sensitivity, $\Delta T_{2\times}$	1.5 °C	1	353.4	10.03	3,047	9,028	702
	4.5 °C	1	356.6	12.33	3,229	9,142	871
Emission/GDP scenario	IS92c	1	345.2	22.16	4,010	8,934	1,520
	IS92e	1	399.2	8.01	2,979	10,272	565
Maximum‡		1	403.6	49.69	5,154	10,507	3,664
Minimum‡		1	296.7	2.92	1,895	7,286	206
Emission year	2005	1.013	364.0	6.78	3,438	9,423	1,003
	2015	1.007	373.5	3.96	3,739	9,779	1,228
Atmospheric lifetime§ (yr)		50–200	120	14.5 ± 2.5	50 ± 5	102	13.3
GWP§, integraton horizon	20 yr	1	290	62.0	5,000	7,900	4,300
	100 yr	1	320	24.5	4,000	8,500	1,700
	500 yr	1	180	7.5	1,400	4,200	520

* Middle case *r*, 3% yr⁻¹; γ 2; $\Delta T_{2\times}$, 2.5 °C; emission/GDP scenario, IS92a; emission year, 1995. Other cases: one parameter set as specified; all others equal to their middle-case values.

† Defined by equation (4); $\chi = 0.1$, $\Delta T_m = 4.25$ °C.

‡ Maximum and minimum values over 81 combinations of *r* = 1, 3, 5% yr⁻¹; γ = 1, 2, 3; $\Delta T_{2\times}$ = 1.5, 2.5, 4.5 °C; emission/GDP scenario, IS92a, c, e.

§ GWP and atmospheric lifetimes from ref. 3. No single lifetime for CO₂ can be defined because of the different rates of uptake by multiple sink processes.

fertilization were included, α would not cancel and information about the level as well as the shape of damages would be required.) For the shape we consider simple geometric functions $D_\gamma[\Delta T(t)] = [\Delta T(t)]^\gamma$ with $\gamma = 1, 2$, and 3 and a more nonlinear 'hockey-stick' damage function (A. S. Manne, personal communication)

$$D_h[\Delta T(t)] = 1 - \left[1 - \left(\frac{\Delta T(t)}{\Delta T_m} \right)^2 \right]^\chi \quad (4)$$

where ΔT_m is a catastrophic level of ΔT and χ is a parameter between 0 and 1. For $\chi = 1$, D_h reduces to the quadratic damage function D_2 ; as χ approaches 0, D_h becomes increasingly convex with damages equal to 100% of GDP when ΔT reaches ΔT_m (D_h is undefined for $\Delta T > \Delta T_m$). Damage functions incorporating additional climate dimensions (for example, meridional temperature gradients and precipitation) could be substituted.

EDI depends on future GHG concentrations, climate sensitivity $\Delta T_{2\times}$ (the equilibrium change in ΔT accompanying a doubling of the pre-industrial level of atmospheric CO₂), and the date of the incremental emission. We consider three GHG-emission/GDP scenarios (IS92a, c, and e; these are IPCC middle, low and high cases, respectively¹⁶) with emissions extrapolated linearly and GDP exponentially at their 2075–2100 rates through our integration horizon of 2200; $\Delta T_{2\times}$ = 1.5, 2.5 and 4.5 °C (refs 17, 18); discount rate *r* = 1, 3 and 5% yr⁻¹; and incremental emissions in 1995, 2005 and 2015.

We calculate EDI by comparing scenarios with and without incremental GHG emissions using the Integrated Science Assessment Model¹⁹. The carbon-cycle component^{19,20} is representative of current global carbon-cycle models^{21,22}. It includes an upwelling-diffusion ocean, six-box biosphere, temperature-dependent buffer factor, and logarithmic CO₂ fertilization²³.

Atmospheric N₂O and halocarbon concentrations are calculated using a mass-balance model with removal rates inversely proportional to the atmospheric lifetimes. Tropospheric CH₄ is calculated by simulating the main chemical processes influencing global levels of CH₄, CO and OH radicals. CH₄ and CO removal rates take account of oxidation by OH radicals, soil uptake and stratospheric transport. Calculated OH radical concentrations depend on CH₄, CO, NO_x, non-methane hydrocarbons,

tropospheric O₃, and H₂O concentrations. Indirect radiative effects of CH₄ and halocarbons due to their influence on H₂O and O₃ are not incorporated.

Global-mean surface and ocean temperatures are calculated using an upwelling-diffusion climate model with ocean-mixing parametrization consistent with the model used by IPCC^{18,24}. In addition to $\Delta T_{2\times}$, the advection velocity, diffusive coefficient and polar-temperature feedback parameter determine the simulated rate of ocean heat transfer.

Illustrative values of EDI are given in Table 1. Compared with the GWP with 100-yr horizon, the 'middle case' ($\Delta T_{2\times}$ = 2.5 °C, emission/GDP scenario IS92a, γ = 2, *r* = 3% yr⁻¹, and emission year 1995) EDI is ~10% larger for long-lived N₂O and CFC-12; 55% smaller for short-lived CH₄ and HCFC-22, and 20% smaller for intermediate-lived CFC-11.

The sensitivity of EDI to parameter values is inversely correlated with atmospheric residence time; over all 81 combinations of *r*, γ , $\Delta T_{2\times}$, and emission/GDP scenario considered, the ratios of the maximum to minimum EDI values are 1.36, 1.44, 2.72, 17.0 and 17.8 for N₂O, CFC-12, CFC-11, CH₄ and HCFC-22, respectively. The relative sensitivity to parameters also varies with residence time: CH₄, HCFC-22 and CFC-11 are proportionately most sensitive to *r*, γ and the emission/GDP scenario; CFC-12 to *r*, the emission/GDP scenario and γ ; and N₂O to the emission/GDP scenario, *r* and γ . Smaller discount rates yield smaller EDIs for all GHGs considered, as these GHGs have shorter residence times than CO₂ and a smaller discount rate gives greater weight to the distant future. For short-lived CH₄ and HCFC-22, larger γ and higher emission and GDP growth also yield smaller EDIs, as these parameter changes increase the incremental damages more in the distant than the near future. Conversely, higher emission and GDP growth increase the EDI for N₂O and CFC-12, although larger γ decreases them. The EDIs are least sensitive to changes in $\Delta T_{2\times}$, as its effect on incremental damages is more linear than the effect of changes in γ and emission scenario.

To explore further the sensitivity to nonlinearity in damages, we evaluate EDIs using an extreme 'hockey-stick' specification, setting $\chi = 0.1$ and $\Delta T_m = 4.25$ °C (just greater than the maximum $\Delta T(t)$ = 4.24 °C for $\Delta T_{2\times}$ = 2.5 °C and scenario IS92a at our integration horizon 2200). For CH₄ and HCFC-22, the resulting EDI is between the values obtained using γ = 2 and 3, as these

GHGs are largely removed from the atmosphere by the time ΔT reaches values where the 'hockey-stick' damage function is sharply nonlinear. For the longer-lived GHGs, the resulting EDI is modestly smaller than the EDI with $\gamma = 3$.

The timing of GHG emissions is of central interest for policy^{11,25}. The EDI can be used to compare incremental emissions at different dates, for example 2005 and 2015, using CO₂ emissions in 1995 as the base. The EDI for CO₂ increases then decreases with decadal CO₂ emission delays, reflecting the opposing effects of increasing the incremental forcing in future years when incremental damages are larger but discounting is greater. EDIs for CH₄ decrease with emission year as emissions are delayed; those for the other GHGs considered increase with emission year. These results suggest that the welfare loss from climate change associated with IS92a emissions could be reduced by incrementally reallocating CO₂ emissions from 2005 to 1995 and 2015, delaying CH₄ emissions from 1995, and advancing emissions of N₂O, CFC-11, CFC-12 and HCFC-22 from 2005 and 2015 to 1995.

As the effects of climate change may depend more on its rate than its magnitude, damage functions based on $d[\Delta T(t)]/dt$ have been proposed²⁶. Application of such functions to EDI is problematical, because the effect of an incremental emission is to first increase then decrease $d[\Delta T(t)]/dt$ as a function of t ; for typical parameter values, the integrated incremental damages are negative. Viewing EDI as a function of the integration horizon, it starts positive, turns negative if either numerator or denominator changes sign before the other, then turns positive when both terms become less than zero (with a singularity when the denominator equals zero). For long integration horizons, EDIs calculated using a damage function equal to $\{d[\Delta T(t)]/dt\}^\gamma$ typically approximate the corresponding EDIs using D_γ .

Neither the GWP nor the EDI as defined here incorporate the effect of uncertainties on policy choice. Because information about climate change, its consequences, and mitigation approaches will change over time, near-term policies should be designed to accommodate sequential change¹¹. With uncertainty, emissions of long-lived GHGs like CO₂ are more costly than emissions of shorter-lived GHGs like CH₄, because they constrain the set of achievable future climate states more severely. It may be useful to characterize uncertainty about $\Delta T_{2\times}$ and other parameters using probability distributions, and to incorporate the economic value of the option to decrease future radiative forcing that is lost by emitting longer-lived GHGs in an extended EDI.

Values of EDI are broadly comparable in magnitude and in range of uncertainty to GWP. Nevertheless, EDI offers significant advantages for policy analysis. Because EDI is defined in terms of economic welfare, it is directly relevant to policy choice; moreover, questions about the damage function and discount rate are clearly defined and may be addressed by research on climate impacts and preferences for impacts at different dates. EDI can also incorporate non-climate effects of GHG emissions on welfare, such as CO₂ fertilization and stratospheric-ozone depletion. In contrast, GWP does not measure welfare changes, non-climate effects cannot be incorporated, and the appropriate time horizon cannot be determined by analysis⁴⁻⁷. Uncertainty about the appropriate values of EDI may be an accurate reflection of uncertainty about the consequences for welfare of climate change, rather than a limitation of the index. □

Received 13 October 1995; accepted 17 April 1996.

- Shine, K. P., Derwent, R. G., Wuebbles, D. J. & Morcrette, J.-J. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 41–68 (Cambridge Univ. Press, 1990).
- Isaksen, I. S. A., Ramaswamy, V., Rodhe, H. & Wigley, T. M. L. in *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 47–67 (Cambridge Univ. Press, 1992).
- Albritton, D. L., Derwent, R. G., Isaksen, I. S. A., Lal, M. & Wuebbles, D. J. in *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emissions Scenarios* (eds Houghton, J. T. et al.) 205–231 (Cambridge Univ. Press, 1995).
- Eckaus, R. S. *Energy J.* **13**, 25–35 (1992).
- Reilly, J. M. & Richards, K. R. *Environ. Resour. Econ.* **3**, 41–61 (1993).
- Schmalensee, R. *Energy J.* **14**, 245–255 (1993).
- Kandlikar, M. *Energy Policy* **23**, 879–883 (1995).
- Cropper, M. L. & Oates, W. E. *J. econ. Literat.* **30**, 675–740 (1992).
- Victor, D. G. *Policy Sci.* **24**, 199–222 (1991).
- Lashof, D. A. & Ahuja, D. R. *Nature* **344**, 529–531 (1990).
- Hammit, J. K., Lempert, R. J. & Schlesinger, M. E. *Nature* **357**, 315–318 (1992).
- Wuebbles, D. J., Jain, A. K., Patten, K. O. & Grant, K. E. *Clim. Change* **29**, 265–297 (1995).
- Nordhaus, W. D. *Science* **258**, 1315–1319 (1992).
- Peck, S. C. & Teisberg, T. J. *Energy J.* **13**, 55–77 (1992).
- Manne, A., Mendelsohn, R. & Richels, R. *Energy Policy* **23**, 17–34 (1995).
- Leggett, J., Pepper, W. & Swart, R. in *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 69–95 (Cambridge Univ. Press, 1992).
- Mitchell, J. F. B., Manabe, S., Meleshko, V. & Tokioka, T. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 131–172 (Cambridge Univ. Press, 1990).
- Gates, W. L., Mitchell, J. F. B., Boer, G. J., Cubash, U. & Meleshko, V. P. in *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 97–134 (Cambridge Univ. Press, 1992).
- Jain, A. K., Kheshgi, H. S. & Wuebbles, D. J. *Integrated Science Model for Assessment of Climate Change* (UCRL-JC 116526; Lawrence Livermore National Lab., Livermore, CA, 1994).
- Jain, A. K., Kheshgi, H. S., Hoffert, M. I. & Wuebbles, D. J. *Global biogeochem. Cycles* **9**, 153–166 (1995).
- Schimmel, D. S. et al. in *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emissions Scenarios* (eds Houghton, J. T. et al.) 35–71 (Cambridge Univ. Press, 1995).
- Eting, I. G., Wigley, T. M. L. & Heimann, M. *Future Emissions and Concentrations of Carbon Dioxide: Key Ocean/Atmosphere/Land Analyses* (Division of Atmospheric Res., CSIRO, Australia, 1995).
- Bacastow, R. & Keeling, C. D. in *Carbon in the Biosphere* (eds Woodwell, G. M. & Pecan, E. V.) 86–135 (NTIS, Springfield, Virginia, 1973).
- Bretherton, F. P., Bryan, K. & Woods, J. D. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 173–193 (Cambridge Univ. Press, 1990).
- Schlesinger, M. E. & Jiang, X. *Nature* **350**, 219–221 (1991).
- Peck, S. C. & Teisberg, T. J. *Clim. Change* **28**, 289–315 (1994).

ACKNOWLEDGEMENTS. This work was supported by the US Department of Energy, Office of Health and Environmental Research.

Occurrence patterns of foreshocks to large earthquakes in the western United States

Rachel E. Abercrombie*† & Jim Mori‡

* Department of Earth Sciences, University of Southern California, Los Angeles, California, 90089, USA

‡ US Geological Survey, Pasadena, California, USA

OBSERVATIONS of foreshocks preceding large earthquakes provide one of the few well documented cases of premonitory events that are clearly related to a subsequent earthquake. Unfortunately, the apparent randomness of foreshock occurrence—they precede some events and not others—has severely hampered their use in reliable earthquake prediction. Understanding the factors that control foreshock occurrence is critical for determining how large earthquakes initiate and whether reliable short-term prediction will ever be possible¹. Here we report the results of a comprehensive study of the occurrence patterns of foreshocks to large earthquakes in the western United States. The incidence of foreshocks decreases with increasing depth of the mainshock, and also depends on the mainshock slip orientation. This pattern of occurrence may be explained by a decrease in small-scale crustal heterogeneity with increasing depth, and suggests that increasing normal stress (both regional tectonic stress and lithostatic load) inhibits the occurrence of foreshocks. No relationship is observed between any aspect of foreshock occurrence and the magnitude of the subsequent mainshock, suggesting that

† Present address: Institute of Geological and Nuclear Sciences, Kelburn Research Centre, 32 Salamanca Road, PO Box 1320, Wellington, New Zealand.

the eventual size of the mainshock may be independent of the earthquake nucleation process, or that foreshocks are not part of this process.

Earthquakes are sometimes preceded by single or multiple events of smaller magnitude, known as foreshocks². How foreshocks are related to the subsequent large earthquake (mainshock) and why they precede some and not others is still unknown. Most theories assume that heterogeneity in the crust contributes to foreshock occurrence². For example, foreshock occurrence has been related to heterogeneities within a zone of accelerating pre-seismic slip, similar to that observed in laboratory experiments^{3,4}. Previous studies of foreshocks have been limited by the data quantity and quality. Attempts to find any relationship between characteristics of the foreshock sequences and the mainshock magnitude have been unsuccessful⁵⁻⁷. In California, a possible decrease in foreshock sequence duration with mainshock depth was noted², as was a greater tendency for more strike-slip earthquakes to have foreshocks than thrusts, but the available data were insufficient to produce a convincing result. In the decade and more since these studies were performed, many earthquakes have been well recorded by the dense seismic networks operating in California and Nevada. Also, more seismometers and enhanced analysis techniques have lead to considerable improvement in resolution of source parameters such as earthquake depth and focal mechanism. This increase in both quantity and quality of available earthquake data now enables us to determine robust relationships governing foreshock occurrence. Here we focus on how foreshock occurrence is related to the hypocentral depth, style of faulting and magnitude of the subsequent mainshock in order to improve our understanding of the earthquake nucleation process.

The characteristics of foreshock occurrence can only be studied reliably in areas and time periods in which regional seismic networks are able to detect small earthquakes and in which many large earthquakes have occurred. The study areas selected are the parts of California and Nevada in which seismic networks

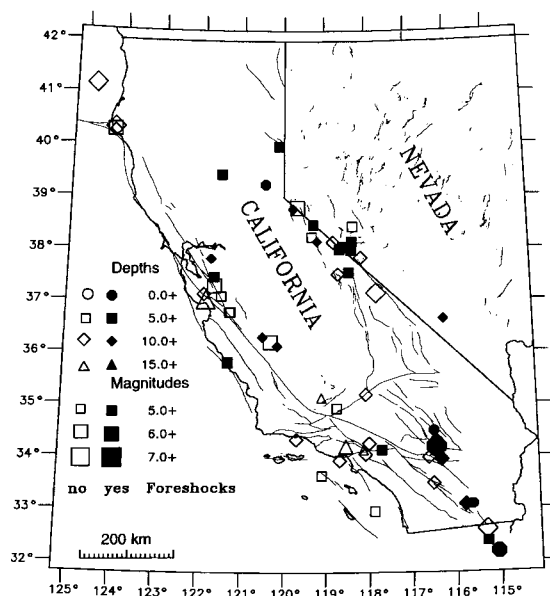


FIG. 1 Map showing the locations of the 59 $M \geq 5$ earthquakes included in this study. The symbols are depth-dependent; filled symbols indicate earthquakes preceded by immediate foreshocks, and open symbols, those without. In the case of multiple earthquake sequences, the first earthquake with $M \geq 5$ is considered the mainshock and later events are not included in this study, as the foreshocks of the later events cannot be distinguished from the aftershocks of the first.

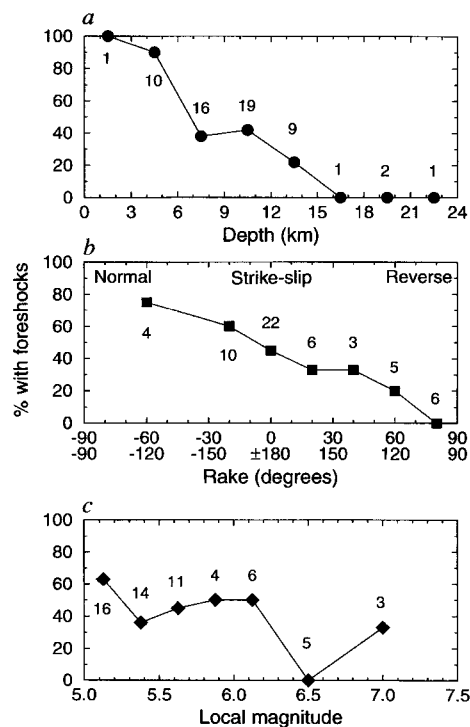


FIG. 2 Foreshock occurrence as a function of mainshock (a) depth, (b) rake, and (c) magnitude. A rake of -90° corresponds to normal faulting, 0° and $\pm 180^\circ$ correspond to strike-slip motion, and 90° to reverse or thrust faulting. The data are from Table 1 and the numbers above or below each data point correspond to the total number of mainshocks in that interval. The occurrence pattern of foreshocks is dependent on the depth and style of faulting of the mainshock, but appears independent of the mainshock magnitude. We investigated these relationships using the Student's *t*-test to evaluate the difference between populations⁵⁶. The cumulative sum curve of $1 - p$ (if the event has foreshocks) and $-p$ (no foreshocks) against depth was plotted, where p is the proportion of earthquakes in the whole data set with foreshocks. The earthquakes were then divided into three depth ranges at the two gradient changes in the cumulative sum: ≤ 8.3 km, $8.3-12.1$ km and >12.1 km (80%, 33% and 1% having foreshocks, respectively). We found the shallowest group to be different from the others at the 99% significance level, and the deeper groups significantly different from one another at the 95% confidence level. Similarly we divided the earthquakes into two populations with 'rake' $\leq 20^\circ$ and $>20^\circ$ (52% and 13% having foreshocks, respectively) and these were found to be different at the 99% confidence level. Dividing the data set by magnitude (for example, at magnitude 5.5 or 6) however, produced no statistically significant difference in the populations.

have had complete detection of earthquakes with magnitude $M \geq 2$ since 1975. Excluding aftershocks and swarm activity immediately around the Mammoth Lakes volcanic area, 59 earthquakes with $M \geq 5$ occurred during this period (Table 1, Fig. 1). The network catalogues were searched for immediate foreshocks, $M \geq 2$, occurring within 5 km and 30 days of these large earthquakes. Such small limits in time and space were chosen as these events can most clearly be related to the nucleation of the subsequent large earthquake⁸. Also, tight spatial and temporal limits enable foreshocks to be distinguished easily from background seismicity. In all but one sequence, foreshocks continued to within 24 hours of the mainshock.

There is no clear geographical separation of earthquakes with and without foreshocks, although a slightly higher proportion of earthquakes in the Mojave shear zone and along the California-Nevada border have foreshocks than along the San Andreas fault zone. We therefore analyse the data set as a whole and find that 44% of the earthquakes in the whole data set had foreshocks,

TABLE 1 Hypocentral parameters of the $M \geq 5$ earthquakes considered in this study

Date (Yr/month/d h:min:s)	Name	Lat. (°)	Long. (°)	Depth (km)	M _L	Strike (°)	Dip (°)	Rake (°)	N	Duration (h)	
1975/6/1	01:38:49	Galway Lake ²¹	34.52	116.50	4.5*	5.0	160	70	180	6	720
1975/8/1	20:20:05	Oroville ²²	39.45	121.54	5.5	5.7	180	65	-70	11	720
1977/2/22	06:24:06	Bridgeport Is. ^{TS}	38.49	119.29	6.7*	5.0				1	116.98
1978/8/13	22:54:53	Santa Barbara ²³	34.35	119.70	12.8*	5.1	280	27	57	-	-
1978/9/4	21:54:52	Diamond Valley ²⁴	38.79	119.79	10.1*	5.2	315	80	160	9	34.483
1978/10/4	16:42:48	Wheeler Crest ²⁵	37.53	118.70	14.9	5.4	015	90	10	-	-
1979/1/1	23:14:39	New Year/Malibu ²⁶	33.94	118.68	11.3*	5.2	280	60	85	-	-
1979/2/22	15:57:29	Doyle ²⁷	39.99	120.12	8.2*	5.3	40	70	-20 (-160)	1	8.67
1979/3/15	21:07:16	Homestead Valley ²⁸	34.33	116.44	2.5*	5.3		180		19	720
1979/8/6	17:05:22	Coyote Lake ²⁹	37.10	121.51	8.9*	5.8	024	80	180	-	-
1979/10/7	20:54:40	Bridgeport II ³⁰	38.23	119.35	8.4*	5.0	150	90	165 (0)	-	-
1979/10/15	23:16:53	Imperial Valley ³¹	32.61	115.32	12.3*	6.4	323	90	180	-	-
1980/1/24	19:00:08	Livermore Valley ³²	37.83	121.77	11.8	5.8	167	85	-160	1	0.0242
1980/2/25	10:47:38	Anza/White Wash ³³	33.50	116.51	12.7	5.5	305	70	-162	-	-
1980/6/9	03:28:19	Victoria ³¹	32.19	115.08	4.6*	6.1	318	90	180	2	480
1980/6/29	07:46:13		38.01	118.66	9.3*	5.0				1	0.0242
1980/9/7	01:30:42	ref. 30	38.07	118.58	8.3*	5.1	74	72	0 (162)	21	43.412
1980/11/8	10:27:30	Eureka ³⁴	41.15	124.75	11.0	6.9	45	90	0	-	-
1980/11/28	18:21:13	ref. 35	39.26	120.46	3.6*	5.1	219	80	10 (170)	2	1.159
1980/12/28	22:58:08	ref. 30	38.16	118.37	6.9*	5.1	248	77	-20 (-15)	5	26.91
1981/4/26	12:09:28	Westmoreland ³⁶	33.10	115.63	3.8*	5.7	~200	~60	~-50 (~-140)	85	120
1981/9/4	15:50:50	Catalina ^{TS}	33.65	119.09	6.0*	5.5	325	90	-152	-	-
1982/9/24	07:40:24	ref. 30	37.85	118.16	10.7*	5.5	248	77	-20 (-166)	-	-
1982/10/25	22:26:03		36.32	120.51	12.1*	5.4				3	1.809
1982/12/28	19:06:24	ref. 30	37.99	118.39	9.1	5.2	162	61	-163 (-30)	2	5.217
1983/5/2	23:42:38	Coalinga ³⁷	36.23	120.32	9.8*	6.7	127	23	90	-	-
1983/8/29	10:10:31	San Simeon Is. ^{TS}	35.84	121.34	5.9*	5.2	311	51	133	3	21.74
1984/4/24	21:15:18	Morgan Hill ³⁸	37.31	121.68	8.6*	6.2	325	90	175	-	-
1985/1/24	11:27:21	ref. 30	38.15	118.82	13.3*	5.3	202	80	0 (-170)	-	-
1985/8/4	12:01:55	Kettleman Hills ³⁹	36.14	120.15	10.1	5.6	145	25	110	6	22.079
1986/1/26	19:20:50	TS	36.80	121.28	8.9*	5.5	350	90	165	-	-
1986/3/31	11:55:39	Mount Lewis ⁴⁰	37.48	121.68	8.1	5.7	355	80	175	3	178
1986/7/8	09:20:44	North Palm Springs ⁴¹	34.00	116.61	11.4	5.6	300	45	180	-	-
1986/7/13	13:47:08	Oceanside ⁴²	32.97	117.87	8.8	5.4	110	50	70	-	-
1986/7/20	14:29:45	Chalfant ^{KS}	37.57	118.44	6.7*	5.9	220	65	-10	13	168
1987/2/7	03:45:14	North Cerro Prieto ^{TS}	32.39	115.31	6.0*	5.4		180		2	15.77
1987/10/1	14:42:20	Whittier ⁴³	34.06	118.09	14.6	5.9	270	25	90	-	-
1987/11/24	01:54:14	Elmore Ranch ^{44,45}	33.09	115.79	10.6	6.2	217	79	4	2	0.357
1988/2/20	08:39:57	TS	36.79	121.31	9.7*	5.1	135	50	10	-	-
1988/6/10	23:06:43	Gorman ^{TS}	34.94	118.74	6.8*	5.4	155	65	150	-	-
1988/9/19	02:56:31	Garfield Hills ³⁰	38.46	118.34	9.0	5.3	138	85	-175 (-5)	-	-
1988/12/3	11:38:26	Pasadena ⁴⁶	34.15	118.13	15.5	5.0	245	70	0	-	-
1989/8/8	08:13:27	Lake Elsmar ^{TS}	37.14	121.93	13.9*	5.4	301	63	152	-	-
1989/10/18	00:04:15	Loma Prieta ^{47,48}	37.04	121.88	19.0	7.0	130	70	135	-	-
1990/2/28	23:43:36	Upland ⁴⁹	34.14	117.69	5.5	5.4	310	70	0	2	3.104
1990/10/24	06:15:19	Lee Vining ^{30,50}	38.05	119.12	11.7	5.6	55	60	-10	1	0.0022
1991/6/28	14:43:54	Sierra Madre ⁵¹	34.26	118.00	12.0	5.4	242	50	90	-	-
1991/8/17	19:29:40	Honeydew ⁵²	40.29	124.24	8.5*	6.0	311	22	51	-	-
1991/9/17	21:10:29	San Simeon II ^{MP}	35.82	121.33	8.0	5.2	297	64	113	-	-
1992/3/8	03:43:04	Petrolia ⁵²	40.26	124.23	10.7*	5.3	120	90	180	-	-
1992/4/23	04:50:23	Joshua Tree ⁵³	33.96	116.32	14.4	6.1	340	90	160	5	2.415
1992/4/25	18:06:05	Petrolia ⁵⁴	40.33	124.23	10.6	6.5	350	13	106	-	-
1992/6/28	11:57:34	Landers ⁵³	34.20	116.44	4.5	7.3	350	90	170	19	720
1992/6/29	10:14:20	Little Skull Mtn ^{KS}	36.72	116.31	11.8	5.2	60	70	-70	3	9.760
1992/7/11	18:14:16	Mojave ^{TS}	35.21	118.07	10.7*	5.7	20	80	30	-	-
1993/5/17	23:20:49	Eureka Valley ^{KS}	37.17	117.79	11.9	6.1	174	32	-126	-	-
1993/5/28	04:47:40	Wheeler Ridge ^{TS}	35.15	119.10	21.4*	5.2		~180		-	-
1994/1/17	12:30:55	Northridge ⁵⁵	34.21	118.54	18.7	6.7	105	35	100	-	-
1994/9/12	12:23:44	Double Spring Flat ^{KS}	38.83	119.66	8.0*	6.3	21	81	-25	-	-

N is the number of foreshocks with $M \geq 2$ occurring within 5 km horizontally and 30 days of the mainshock, and 'duration' is the duration of the foreshock sequence ($M \geq 2$). Earthquake names are included where assigned by previous authors. Focal mechanisms and hypocentral depths of the mainshocks are taken from detailed previous studies where available. First motion focal mechanisms are preferred as the most relevant to the nucleation process. An asterisk following the depth indicates that the catalogue depth is used. If the fault plane is unknown, the rake of the other plane is given in brackets. Focal mechanisms determined in this study are calculated as in ref. 37 using phase picks from the Californian catalogues. TS, this study; KS, K. Smith, personal communication; MP, M. Pasyanos, personal communication. The depths and focal mechanisms of the foreshocks are not considered as most are relatively poorly determined. The mean foreshock depths per sequence are highly correlated with the mainshock depths, however, (correlation coefficient, 0.63; significant at the 99% confidence level). Also, in detailed analyses of well recorded foreshock sequences, the depths and focal mechanisms of the foreshocks are very similar to those of the mainshock^{15,40,44,49}.

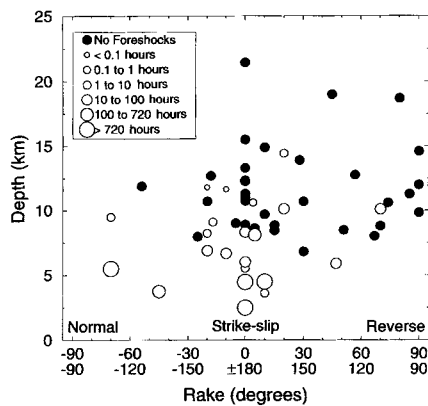


FIG. 3 Foreshock sequence occurrence and duration as a function of mainshock depth and slip orientation. The data are again taken from Table 1. In this dataset, the mean depth of reverse faulting earthquakes is slightly (but not significantly) greater than for normal faulting ones. Within the range -30° to 30° rake, where the depth range is constant, a decrease in foreshock occurrence with increasing rake is observed, so this relationship is not simply an artefact of the depth distribution.

similar to the number found in previous studies in the region^{5,7,9}. It is worth noting, however, that foreshocks preceded only three earthquakes near the conurbations of Los Angeles and the Bay Area, suggesting that they may be of less use for urban hazard warnings in these areas than would be predicted from consideration of the whole data set.

Foreshock occurrence is plotted in Fig. 2 as a function of depth (Fig. 2a), slip orientation (Fig. 2b) and mainshock magnitude (Fig. 2c). Foreshock occurrence clearly decreases with increasing depth and is also dependent on the style of faulting. Most normal faulting earthquakes have foreshocks, compared to about half the strike-slip events and only very few of the reverse faulting ones. Both the relationships between foreshock occurrence and depth and rake are over 95% significant (Fig. 2). No relationship is observed, however, between the mainshock magnitude and the pattern of foreshock occurrence.

Normal stress increases both with depth and also from normal to reverse faulting environments¹⁰, so it is plausible that foreshock occurrence is inhibited by increasing normal stress. Also, in Fig. 3, a tendency for the duration of foreshock sequences to decrease with increasing depth is seen; the correlation coefficient between the logarithm (base 10) of the duration of the foreshock sequence and mainshock depth is 0.61, 99% significant. This has previously been suggested from a smaller data set⁷. This observation supports the hypothesis that increasing normal stress (both regional tectonic stress and loading stress with depth) inhibits foreshock occurrence, and it seems that at high normal stress foreshock sequences tend to be short or non-existent.

Foreshocks, like other small earthquakes, can be interpreted as the breaking of small patches of fault as they reach a critical stress, but which do not propagate further to become large earthquakes. It would seem plausible, therefore, that more small earthquakes, and similarly more foreshocks, would occur in regions with a higher degree of small-scale heterogeneity. The idea of heterogeneity decreasing with depth and normal stress is not new, and is consistent with the frequency-magnitude statistics of earthquake occurrence in California; the ratio of small ($M = 2$) to larger earthquakes is greater at shallower depths than at deeper ones¹¹. This result suggests that, in general, a dynamic rupture nucleating at depth is less likely to be stopped by small-scale heterogeneity—and more likely to grow into a larger earthquake—than a similar initiation at more shallow depths. This pattern is consistent with the observed depth-dependence of foreshock occurrence. Heterogeneity within fault zones affecting earthquake rupture could be compositional, or geometric, or related to spatial variations in

fault gouge properties such as fluid pressures, or even aseismic creep. All of these are expected to vary with depth and normal stress, and discriminating amongst them will be difficult.

Laboratory studies of frictional sliding observe pre-seismic slip over a critical slip distance D_c before unstable dynamic failure^{3,4}. D_c is observed to decrease in size with increasing stress and shear localization¹² and, therefore, may be expected to decrease with depth¹³, and so be related to our foreshock observations. It is unclear, however, whether there is a direct connection, as D_c is estimated to have dimensions considerably smaller than the volumes in which foreshocks occur¹⁴. There is also no apparent reason why a change in D_c should inhibit foreshock occurrence. We therefore prefer the hypothesis of decreasing small-scale heterogeneity with depth, which also explains the observed depth dependence of the earthquake frequency-magnitude distribution.

Perhaps the most important question concerning earthquake nucleation is whether the magnitude of an earthquake is controlled primarily by its nucleation process (for example, the amount and area of pre-seismic slip), or whether all earthquakes initiate in a similar manner, and the final slip and rupture area depend on the local stress state¹ and on the dynamics of the earthquake rupture itself. The first, 'nucleation-controlled', scenario is the most optimistic for short-term prediction; the second, 'rupture-controlled' scenario implies that the magnitude of the earthquake is unknown until the rupture propagation stops. The role of foreshocks in the earthquake initiation process is still unclear. In the 'nucleation-controlled' model they could be interpreted as part of the nucleation process of a large event, probably caused by pre-seismic slip^{4,15}. In the 'rupture-controlled' case, foreshocks are simply small earthquakes, easily stopped by the local heterogeneity, in the hypocentral region of the mainshock. The eventual mainshock is a rupture initiation which is not stopped and cascades into a larger earthquake. In this 'rupture-controlled' model, most small earthquakes will not evolve into large earthquakes, and foreshocks may not be different from other short sequences of smaller earthquakes which do not culminate in a larger event.¹⁶

Recent studies of the onset of large earthquakes^{17,18} have found that they commonly initiate with a small subevent, <1% of the total seismic moment. These small initial subevents can also be interpreted in terms of the two models of earthquake initiation outlined above. In the 'rupture-controlled' model, they can be considered as ordinary small earthquakes (perhaps the last foreshocks?) which dynamically trigger a neighbouring large subevent¹⁷. Alternatively, in the 'nucleation-controlled' model, these initial subevents may be considered part of a nucleation phase with properties differing from a dynamic rupture, and whose size is related to that of the final earthquake¹⁸.

No pre-seismic slip of the kind seen in laboratory experiments of frictional sliding has ever been unambiguously observed before earthquakes, however¹⁹, indicating that if it does occur in the Earth, it must be very small. Some attempts to extrapolate the laboratory results to real crustal faults have proposed that pre-seismic slip could be related to the duration and volume of the foreshock sequence⁴. These models predict a correlation between the mainshock size and the spatial and temporal extent of the foreshock sequence, which can be investigated with the data collected in the present work. Other models, which do not imply size dependence^{14,20}, are more difficult to evaluate.

In the present study, no relationship is seen between the mainshock magnitude and foreshock occurrence (Fig. 2). We also find no relationship between mainshock magnitude and the logarithm (base 10) of the duration of the foreshock sequence (correlation coefficient, 0.17; 95% confidence limits -0.23 to 0.52), between the mainshock magnitude and the number of $M \geq 2$ foreshocks (correlation coefficient, 0.16; 95% confidence limits -0.24 to 0.51), or between mainshock magnitude and the magnitude of the largest foreshock (correlation coefficient, 0.17; 95% confidence limits -0.23 to 0.52). This is consistent with

previous studies⁵⁻⁷ which also found no relationship between foreshock occurrence and the magnitude of the subsequent mainshock. A relationship between the foreshock sequence and the subsequent mainshock magnitude is not explicitly required by the 'nucleation-controlled' model, although it has been suggested⁴, and might be expected in a model in which the rupture is controlled by the nucleation process. The lack of a relationship between foreshock occurrence and mainshock magnitude observed here is therefore more consistent with the second ('rupture-controlled') model. □

Received 24 October 1995; accepted 16 April 1996.

1. Brune, J. N. *J. geophys. Res.* **84**, 2195-2198 (1979).
2. Mogi, K. *Bull. Earthquake Res. Inst.* **41**, 615-658 (1963).
3. Dieterich, J. H. *J. geophys. Res.* **84**, 2161-2168 (1979).
4. Ohnaka, M. *Tectonophysics* **211**, 149-178 (1992).
5. Jones, L. M. *Bull. seism. Soc. Am.* **74**, 1361-1380 (1984).
6. Jones, L. M. & Molnar, P. *J. geophys. Res.* **84**, 3596-3608 (1979).
7. Doser, D. I. *Bull. seism. Soc. Am.* **80**, 110-128 (1989).
8. Jones, L. M. *Bull. seism. Soc. Am.* **75**, 1669-1679 (1985).
9. Savage, M. K. & dePollo, D. M. *Bull. seism. Soc. Am.* **83**, 1910-1938 (1993).
10. Sibson, R. H. *Nature* **249**, 542-544 (1974).
11. Mori, J. & Abercrombie, R. E. *Eos (suppl.)* **76**, F388 (1995).
12. Marone, C. & Kilgore, B. *Nature* **362**, 618-621 (1993).
13. Scholz, C. H. *Nature* **336**, 761-763 (1988).
14. Dieterich, J. H. in *Earthquake Source Mechanics* (eds Das, S., Boatwright, J. & Scholz, C.) 37-47 (Am. Geophys. Un., Washington DC, 1986).
15. Dodge, D. A., Beroza, G. C. & Ellsworth, W. L. *J. geophys. Res.* **100**, 9865-9880 (1995).
16. Jones, L. M., Console, R., Di Luccio, F. & Murro, M. *Eos (suppl.)* **76**, F388 (1995).
17. Abercrombie, R. E. & Mori, J. *Bull. seism. Soc. Am.* **84**, 725-734 (1994).
18. Ellsworth, W. L. & Beroza, G. C. *Science* **268**, 851-855 (1995).
19. Wyatt, F. K., Agnew, D. C. & Gladwin, M. *Bull. seism. Soc. Am.* **84**, 768-779 (1994).
20. Dieterich, J. H. *Tectonophysics* **211**, 115-134 (1992).
21. Kanamori, H. & Fuis, G. *Bull. seism. Soc. Am.* **66**, 2017-2037 (1976).
22. Langston, C. A. & Butler, R. *Bull. seism. Soc. Am.* **66**, 1111-1120 (1976).
23. Corbett, E. J. & Johnson, C. E. *Bull. seism. Soc. Am.* **72**, 2201-2226 (1982).
24. Somerville, M. R., Peppin, W. A. & VanWormer, J. D. *Bull. seism. Soc. Am.* **70**, 1547-1555 (1980).
25. Peppin, W. A., Honjas, W., Sommerville, M. R. & Vetter, U. R. *Bull. seism. Soc. Am.* **79**, 67-76 (1989).

26. Hauksson, E. & Salvidar, G. V. *J. geophys. Res.* **94**, 9591-9606 (1989).
27. VanWormer, J. D. & Ryall, A. S. *Bull. seism. Soc. Am.* **70**, 1557-1572 (1980).
28. Hutton, L. K. *et al. Calif. Geol.* **33**, 110-114 (1980).
29. Liu, H.-L. & Helmbarger, D. V. *Bull. seism. Soc. Am.* **73**, 201-218 (1983).
30. dePollo, C. M., Peppin, W. A. & Johnson, P. A. *Tectonophysics* **225**, 271-299 (1993).
31. Silva, P. & Masuda, T. *J. geophys. Res.* **90**, 7639-7651 (1985).
32. Bolt, B. A., McEvilly, T. V. & Urhammer, R. A. *Bull. seism. Soc. Am.* **71**, 451-463 (1981).
33. Given, J. W. thesis, California Inst. Technol. (1983).
34. Ekström, G. & Dziewonski, A. M. *Bull. seism. Soc. Am.* **75**, 23-39 (1985).
35. Vetter, U. R. & Ryall, A. S. *J. geophys. Res.* **88**, 8237-8250 (1983).
36. Frez, J. & Gonzalez, J. J. *The Gulf and Peninsular Provinces of the Californias* (Memoir 47, Am. Assoc. Petrol. Geol., Tulsa, Oklahoma, 1991).
37. Eaton, J. P. *Prof. Pap. U.S. geol. Surv.* **1487**, 113-170 (1990).
38. Oppenheimer, D. H., Reasenber, P. A. & Simpson, R. W. *J. geophys. Res.* **93**, 9007-9026 (1988).
39. Ekström, G., Stein, R. S., Eaton, J. P. & Eberhart-Phillips, D. J. *geophys. Res.* **97**, 4843-4864 (1992).
40. Zhou, Y., McNally, K. C. & Lay, T. *Phys. Earth planet. Inter.* **75**, 267-288 (1993).
41. Jones, L. M., Hutton, L. K., Given, D. G. & Allen, C. R. *Bull. seism. Soc. Am.* **76**, 1830-1837 (1986).
42. Hauksson, E. & Jones, L. M. *Bull. seism. Soc. Am.* **78**, 1885-1906 (1988).
43. Hauksson, E. & Jones, L. M. *J. geophys. Res.* **94**, 9569-9589 (1989).
44. Magistrale, H., Jones, L. M. & Kanamori, H. *Bull. seism. Soc. Am.* **79**, 239-251 (1989).
45. Sipkin, S. A. *Bull. seism. Soc. Am.* **79**, 493-499 (1989).
46. Hauksson, E. *J. geophys. Res.* **95**, 15365-15394 (1990).
47. Wald, D. J., Helmbarger, D. V. & Heaton, T. H. *Bull. seism. Soc. Am.* **81**, 1540-1572 (1991).
48. Beroza, G. C. *Bull. seism. Soc. Am.* **81**, 1603-1621 (1991).
49. Hauksson, E. & Jones, L. M. *J. geophys. Res.* **96**, 8143-8165 (1991).
50. Horton, S. P., dePollo, D. M. & Walter, W. R. *Bull. seism. Soc. Am.* (in the press).
51. Hauksson, E. *Bull. seism. Soc. Am.* **84**, 1058-1074 (1994).
52. Velasco, A. A., Ammon, C. J. & Lay, T. *J. geophys. Res.* **99**, 711-728 (1994).
53. Hauksson, E., Jones, L. M., Hutton, L. K. & Eberhart-Phillips, D. J. *geophys. Res.* **98**, 19835-19858 (1993).
54. Oppenheimer, D. *et al. Science* **261**, 433-438 (1993).
55. Hauksson, E., Jones, L. M. & Hutton, L. K. *J. geophys. Res.* **100**, 12335-12355 (1995).
56. Dixon, W. J. & Massey, F. J. *Introduction to Statistical Analysis* 119 (McGraw-Hill, New York, 1969).

ACKNOWLEDGEMENTS. We thank K. Smith, D. Oppenheimer, D. dePollo and M. Pasyanos for providing data and source parameters, and E. Smith for assistance with the statistical analysis. We also thank J. Vidale, D. Wald, P. Reasenber, Y. Ogata and T. Webb for comments and suggestions which improved the manuscript. Catalogue data were obtained from the Northern California Earthquake Data Center, Southern California Earthquake Center Data Center, and University of Nevada, Reno. We are grateful to the USNSF, the USGS and SCEC for partial funding of this research.

A complete skeleton of the giant South American primate *Protopithecus*

Walter Carl Hartwig* & Castor Cartelle†

* Department of Anthropology, George Washington University, Washington DC, 20052, USA

† Instituto de Geociencias, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

A COMPLETE skeleton of a large-bodied New World monkey has been found in Pleistocene cave deposits in the Brazilian state of Bahia. It demonstrates an unprecedented combination of body size, locomotor and cranial morphology. Skeletal features indicate an animal of approximately 25 kg, more than twice the mass of any living South American monkey. We refer the specimen to *Protopithecus brasiliensis* Lund, 1838, a large Pleistocene primate originally represented by only a proximal femur and distal humerus¹⁻⁴. The skeleton resembles species of two distinct New World monkey lineages. The cranium is modified for an enlarged vocal sac typical of living howler monkeys⁵⁻⁷, and the postcranium includes suspensory and brachiating components of locomotion as seen in living spider and woolly spider monkeys⁸. This skeleton confirms that adaptive diversity in neotropical primates was greater in the recent past, and that current interpretations of how their distinctive adaptations evolved should be revised.

The anatomy of the *Protopithecus* skeleton closely resembles the anatomy of the four largest living New World monkeys, typically of mass 6-12 kg and classified together at the subfamily rank (Atelinae). This New World monkey subfamily can be

separated into a radiation of howler monkeys (tribe Alouattini, genus *Alouatta*) and atelin monkeys (tribe Atelini, genera *Ateles*, *Brachyteles*, *Lagothrix*)⁵. *Alouatta* is unique morphologically because of its unusually large hyo-laryngeal apparatus and the effects of this on the basicranium^{6,7}. The atelin New World monkeys are characterized by relatively long forelimbs compared to hindlimbs, which reflects their adaptation to a suspensory positional repertoire and a brachiating mode of locomotion⁸. Howler monkeys locomote by a deliberate above-branch quadrupedalism^{9,10}, and thus their limb proportions are closer to equal. *Protopithecus* is remarkable because it displays cranial features typical of alouattins, but postcranial features characteristic of atelins.

Order Primates Linnaeus, 1758
Suborder Haplorhini Pocock, 1918
Infraorder Platyrrhini E. Geoffroy, 1812
Family Atelidae Gray, 1825
Subfamily Atelinae Gray, 1825

Protopithecus Lund, 1838

Protopithecus brasiliensis Lund, 1838, emended diagnosis.

Type specimen. Universitets Zoologisk Museum (UZM) 1623, a left proximal femur, and UZM 3530, a right distal humerus⁴ of the same individual.

Hypodigm. The type, and Instituto de Geociencias, Universidade Federal de Minas Gerais (IGC-UFMG) 06, a nearly complete skeleton preserving cranium, mandible, several vertebrae, scapular and innominate fragments, upper and lower limb long bones, carpals, tarsals, metacarpals, metatarsals and phalanges.

Locality. The type specimen was recovered in 1836 in Lapá do Periperi, in the Pleistocene limestone cave network of Lagoa

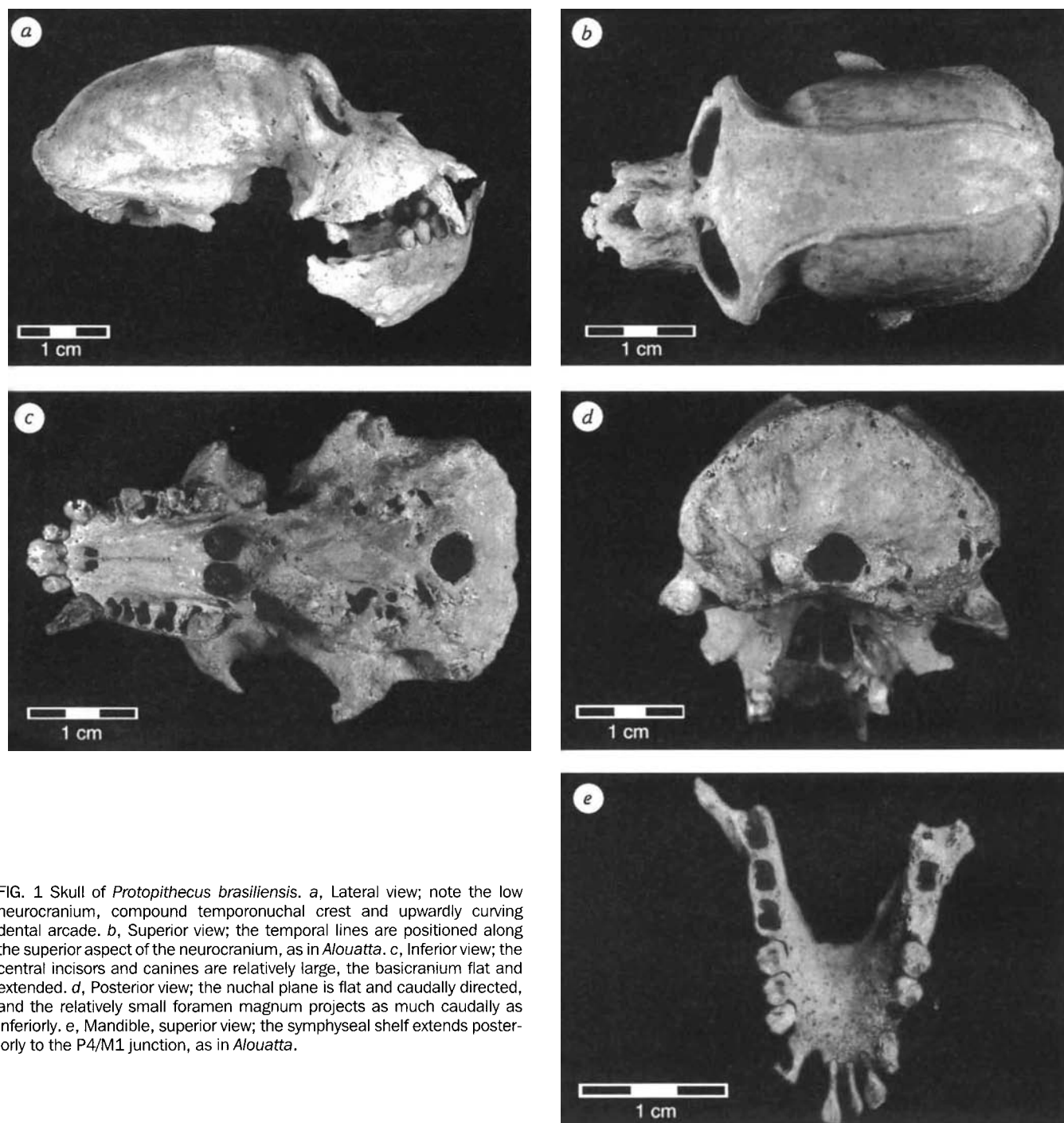


FIG. 1 Skull of *Protopithecus brasiliensis*. *a*, Lateral view; note the low neurocranium, compound temporonuchal crest and upwardly curving dental arcade. *b*, Superior view; the temporal lines are positioned along the superior aspect of the neurocranium, as in *Alouatta*. *c*, Inferior view; the central incisors and canines are relatively large, the basicranium flat and extended. *d*, Posterior view; the nuchal plane is flat and caudally directed, and the relatively small foramen magnum projects as much caudally as inferiorly. *e*, Mandible, superior view; the symphyseal shelf extends posteriorly to the P4/M1 junction, as in *Alouatta*.

Santa, Minas Gerais, Brazil^{4,11}. The complete skeleton described here was recovered in 1992 in Toca da Boa Vista, a large cave near Campo Formosa in Bahia, Brazil. The main entrance to the cave is located at 40°51'39" W longitude and 10°09'36" S latitude at an altitude of 600 m above mean sea level.

Emended diagnosis. *Protopithecus brasiliensis* is a large New World monkey of about 25 kg, characterized by the unique combination of a cranium modified for an enlarged vocal sac and a postcranium modified for a substantial brachiating component of locomotion. No living or extinct primate displays this combination. The cranium has a compound temporonuchal crest, a flat, caudally directed nuchal plane, and an extended basicranium as in *Alouattins*. The *Protopithecus* cranium is distinct from *Alouatta*, however, in lacking dental adaptations to folivory, such as molar shearing crests and reduced incisors. The post-

cranium is extremely robust compared with all other New World monkeys, with strong muscle markings for brachial flexion and leg extension; however, it also has an intermembral index of 1.04, a reduced olecranon process of the ulna, and articular morphology typical of rotational mobility at the glenohumeral joint, closely resembling the morphology of atelin New World monkeys. The *Protopithecus* postcranium is distinct from the atelins *Ateles* and *Lagothrix*, however, by its absolute and relative robustness.

Description. The specimen was discovered *in situ* and is remarkably well preserved (Figs 1 and 2). In all elements of the skeleton, the large size of *Protopithecus* compared with living New World monkeys is evident (Table 1). *Protopithecus* is not a scaled-up version of smaller living atelines, because its limbs are virtually the same length as those of modern species but are approximately 40%

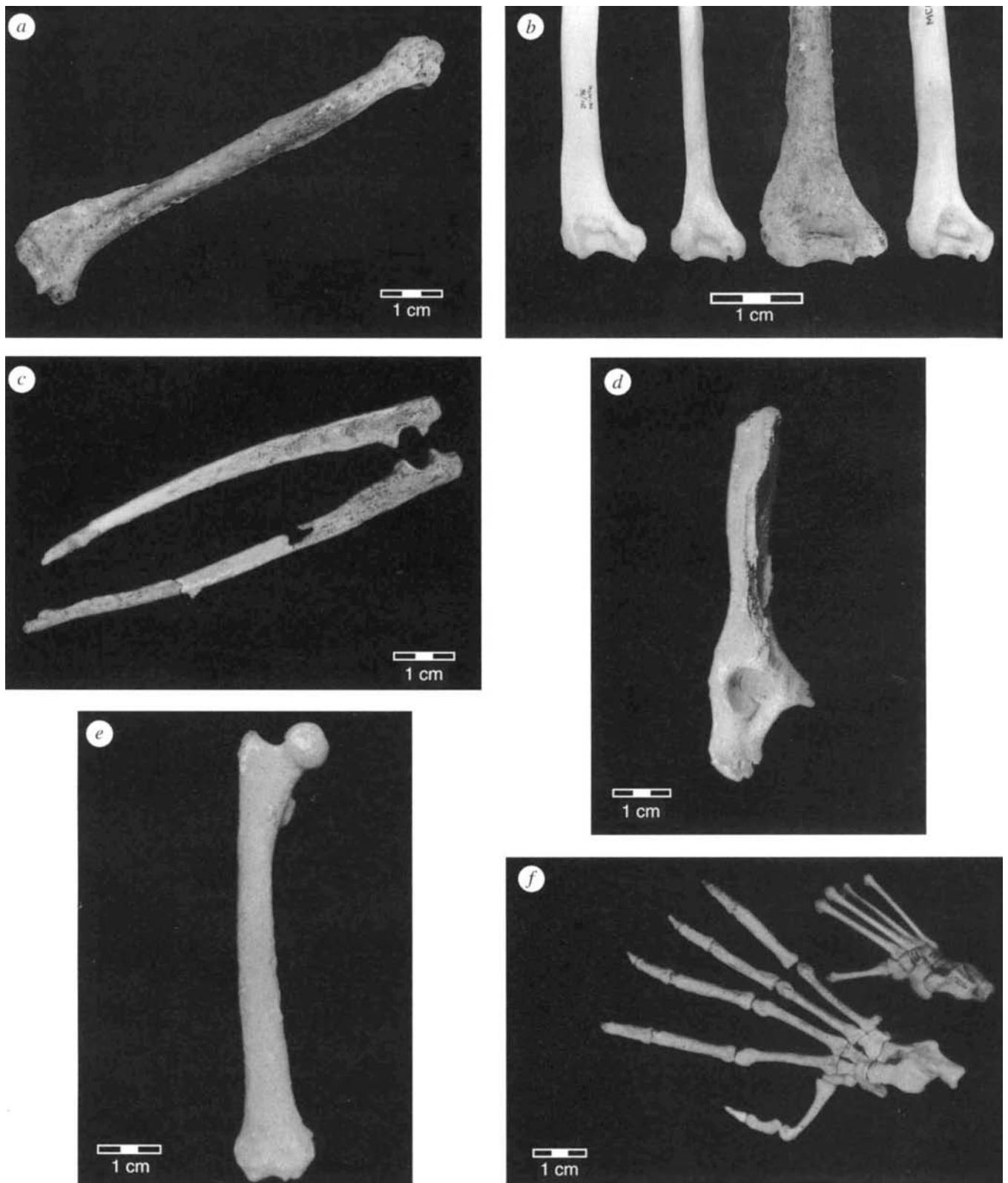


FIG. 2 Selected postcranial elements of *Protopithecus brasiliensis*. *a*, Right humerus, anterior view; note the extremely large brachioradialis flange along the distolateral margin. *b*, Posterior view of distal left humeri of modern atelines and *Protopithecus*; from left to right, *Alouatta*, *Lagothrix*, *Protopithecus*, *Ateles*. *c*, Medial view of left and right ulnae; the short olecranon process and long shaft are characteristic of the spider and woolly

spider monkeys, who display suspensory posture and brachiating locomotion. *d*, Lateral view of right innominate; the ilium is tall and the ischium robust compared to other New World monkeys. *e*, Anterior view of right femur. *f*, Dorsal view of reconstructed right foot of *Protopithecus* (left) and *Ateles paniscus* (right).

more robust. Based on a regression for primate body mass derived from femoral head dimensions ($\log_{10}(\text{femoral head volume}) = 1.196(\log_{10}(\text{bodymass})) + 2.234$; $r = 0.998$, s.e. = 0.028)¹², the skeleton described here reflects a live-animal mass of 24.85 kg, more than twice that of the largest living South American primates¹³. Such a size increase in an arboreal, suspensory primate may be related to a greater dependence on leaves in the diet, as in the case of siamangs versus gibbons¹⁴, or may reflect the general tendency of mammals in the late Pleistocene to become larger bodied¹⁵. The robustness of *Protopithecus* may be a response to competition for dwindling resources brought on by deteriorating climate in the late Pleistocene in this region of Brazil, which today is a relatively dry region bridging Amazonia and the Atlantic coastal forest.

The skull of *Protopithecus* most closely resembles the unique morphological pattern of the howler monkey, which suggests that it also had an enlarged vocal sac. Both have long, low and cylindrical neurocrania, with robust temporal lines that peak above rather than alongside the parietals (Fig. 1a, b). The nuchal plane of *Protopithecus* is flat and faces caudally to a degree that is more similar to *Alouatta* than to any other New World monkey (Fig. 1d). The preserved portion of the mandible is characterized by a deep symphyseal shelf and lower molar alveoli that are shaped and positioned along the base of the ascending ramus as in *Alouatta* (Fig. 1e). For each of these cranial characters, however, the condition in *Protopithecus* is not as pronounced as it is in the living howler monkey.

The appendicular joint morphology and forelimb/hindlimb proportions of *Protopithecus* are virtually identical to those of the modern atelins (Table 2). These New World monkeys have a suspensory posture and brachiating mode of locomotion, and spend much of their time beneath rather than above the branch of support¹⁴. The limb proportions and upper limb rotational mobility of *Protopithecus* suggest similar postural and locomotor adaptations, and contrast with these aspects of howler monkey skeletal anatomy.

Unique features of the *Protopithecus* postcranium include its overall robustness, a well-developed brachioradialis flange on the distal humerus (Fig. 2a), a relatively large and deep iliac fossa for the belly of gluteus medius (Fig. 2d), and a marked gluteal tuberosity on the femoral shaft. These features suggest a relatively heavy New World monkey that relied upon a forelimb that could flex with power and a strong gluteus medius for extension of the hindlimb during locomotion. This combination suggests a pattern of suspensory posture and brachiating locomotion combined with hindlimb-supported climbing. *Protopithecus* may thus have spent more time climbing in and out of the canopy

TABLE 1 Comparison of cranial measurements of *Protopithecus* and four extant New World monkeys

	<i>Protopithecus</i> (N = 1)	<i>Ateles</i> (N = 92)	<i>Brachyteles</i> (N = 11)	<i>Lagothrix</i> (N = 73)	<i>Alouatta</i> (N = 25)
NCL (mm)	110.1	77.9 68.5–84.4	86.6 79.8–91.7	73.7 67.0–81.2	61.2 54.9–68.8
NCB (mm)	72.8	60.8 54.9–65.7	61.9 57.7–65.1	58.6 53.8–63.1	51.0 47.3–56.2
TSL (mm)	150.5	114.1 104.0–122.0	114.8 100.0–122.0	104.9 97.2–114.6	107.6 96.3–121.4
BAS-NAS (mm)	83.4	63.3 55.4–71.9	68.2 58.2–74.4	63.1 57.9–70.0	64.9 57.6–78.2
PL (mm)	43.8	34.4 29.9–40.7	38.7 34.2–44.1	31.6 26.2–37.4	39.9 34.7–59.9
BOB (mm)	70.8	55.6 49.9–64.2	57.3 52.0–61.2	54.0 47.9–59.0	52.3 47.2–60.7

Cranial measurements for *Protopithecus* and the four genera of living ateline New World monkeys are expressed as means and ranges. NCL, neurocranial length; NCB, neurocranial breadth; TSL, total skull length; BAS-NAS, basion-nasion; PL, palate length; BOB, biorbital breadth.

TABLE 2 Comparison of postcranial measurements of *Protopithecus* and four extant New World monkeys

	<i>Protopithecus</i> (N = 1)	<i>Ateles</i> (N = 31)	<i>Brachyteles</i> (N = 3)	<i>Lagothrix</i> (N = 17)	<i>Alouatta</i> (N = 25)
FHD (mm)	25.0	17.9 15.8–20.2	18.2 16.9–19.8	15.0 14.0–15.7	13.4 11.8–15.9
FND (mm)	16.2	9.9 8.0–11.8	10.3 9.5–11.2	8.3 7.4–9.6	7.9 6.4–10.8
FL (mm)	237	205.6 190.5–226.0	202.0 186.5–212.0	166.4 157.5–176.5	154.2 139.0–171.0
BCB (mm)	45.5	31.8 29.1–34.9	29.0 28.0–30.9	27.1 24.2–29.3	23.9 21.4–26.7
HHD (mm)	28.4	20.5 17.8–24.1	19.8 18.2–21.6	20.1 18.6–22.2	19.8 16.9–23.1
HMST (mm)	18.5	11.0 10.0–12.4	10.4 9.4–11.6	10.2 9.2–11.4	9.5 7.3–12.1
BIEPI (mm)	48.0	30.9 28.5–33.2	30.0 26.7–33.0	28.0 25.6–30.0	26.6 22.5–30.8
I-I (mm)	1.04	1.05 1.01–1.07	1.07 1.05–1.08	.98 0.96–1.00	.95 0.92–0.98
FV (ml)	90	40 (N = 1)		20 (N = 1)	20 (N = 1)

Postcranial measurements for *Protopithecus* and the four genera of living ateline New World monkeys are expressed as means and ranges. FHD, femoral head diameter; FND, femoral neck diameter; FL, femoral length; BCB, femoral bicondylar breadth; HHD, humeral head diameter; HMST, humeral midshaft thickness; BIEPI, humeral biepicondylar breadth; I-I, intermembral index (forelimb length/hindlimb length); FV, femoral volume (measured by water displacement in a graduated cylinder).

as part of a more dispersed foraging strategy than that of extant species.

Protopithecus displays distinct morphological features that have been assumed to be highly derived for two different modern lineages. This poses an obvious problem for phylogeny reconstruction in atelines. With reference only to the living forms, either *Protopithecus* is best classified as an atelin and its cranial anatomy regarded as convergent on that of howler monkeys, or, conversely, it is best classified as an alouattin and its postcranial anatomy regarded as convergent on that of atelins. *Protopithecus* may demonstrate that evolutionary changes in cranial and postcranial complexes in atelines were sequential rather than simultaneous. Thus a suspensory/brachiation locomotor adaptation may be the primitive condition for the ateline radiation. Later divergence of howler monkeys may have been predicated on development of an enlarged vocal sac and eventual return to an above-branch quadrupedalism locomotor strategy. In this sense, complete fossil specimens can sometimes reveal a directionality of evolutionary change not always apparent in taxonomically impoverished living faunas. □

Received 1 November 1995; accepted 25 March 1996.

1. Lund, P. W. *Det Kong. Danske Viden. Selsk. Natur. Matem. Aftand.* **8**, 61–144 (1838).
2. Cartelle, C. *Neotrop. Primates* **1**, 8 (1993).
3. Cartelle, C. & Ferreira, M. A. C. *Acta geol. Leopold* **XVII**, 411–414 (1994).
4. Hartwig, W. C. *J. hum. Evol.* **28**, 189–196 (1995).
5. Rosenberger, A. L. & Strier, K. B. *J. hum. Evol.* **18**, 717–750 (1989).
6. Schön, M. A. *Folia primatol.* **15**, 117–132 (1971).
7. Schön, M. A. in *Symposium on Development of the Basicranium* (ed. Bosma, J. F.) 664–677 (U.S. Department of Health Education and Welfare, Bethesda, MD, 1976).
8. Erickson, G. E. *Symp. zool. Soc. Lond.* **10**, 135–164 (1963).
9. Mendel, F. *Folia primatol.* **26**, 36–53 (1976).
10. Schön Ybarra, M. *Am. J. phys. Anthropol.* **63**, 65–76 (1984).
11. Hartwig, W. C. *Hist. Phil. Life Sci.* **17**, 3–16 (1996).
12. Ruff, C. B. in *Body Size in Mammalian Paleobiology: Estimation and Biological Implications* (eds

- Damuth, J. & MacFadden, B. J.) 119–150 (Cambridge Univ. Press, 1990).
13. Peres, C. A. *J. hum. Evol.* **26**, 245–250 (1994).
14. Fleagle, J. G. *Primate Adaptation and Evolution* (Academic, San Diego, 1988).
15. Martin, P. S. & Klein, R. G. *Quaternary Extinctions* (Univ. Arizona Press, Tucson, 1984).

ACKNOWLEDGEMENTS. The fossils of Toca da Boa Vista were discovered by members of Grupo Bambuí de Pesquisas Espeleológicas de Belo Horizonte. We thank M. A. C. Ferreira and Rodrigo Lopes F. for help in the field; A. L. Rosenberger, W. Jungers and M. Schön Ybarra for suggestions; J. Fleagle for support, advice and encouragement; S. Franco, L. Oliveira, G. W. Nunan, R. Thorington, B. Patterson and R. MacPhee for access to comparative specimens; and the Universidade Federal de Minas Gerais and Pontifícia Universidade Católica de Minas Gerais for laboratory and photographic support. This work was supported by the LSB Leakey Foundation.

CORRESPONDENCE and requests for materials should be addressed to C.C. (e-mail: cartelle@csr.ufmg.br).

Predator–prey cycles with period shifts between two- and three-species systems

Michael Begon, Steven M. Sait* & David J. Thompson

Population Biology Research Group, Department of Environmental and Evolutionary Biology, The University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

POPULATION ecology typically focuses on particular species or pairs of species within webs of interacting species to understand variations in their abundance. Classical theory^{1–3} predicts multi-generation cycles in predator and prey abundance. These cycles have received considerable attention^{3–6}, although there is an alternative possibility, of prey generation-length cycles in predator–prey interactions^{7–9}. Does observation of either of these patterns depend on how firmly predators and prey are embedded in their web of interactions? Concurrent investigations of predator–prey dynamics both in isolation and within a larger web have been lacking. Here we report observations of the population dynamics of a simple predator–prey system, and also of a three-species system including predator and prey. The dynamic patterns exhibited by both systems are cyclic, but the increase from two to three species gives rise to a marked shift in cycle period from one to several host generation lengths.

Parasitoids have contributed greatly to our understanding of predator–prey interactions. Parasitoids are insects that attack and usually kill other arthropods when the adult female deposits her eggs in or on preadult hosts, within which the parasitoid larvae develop. Here we have studied the interaction between the ichneumonid parasitoid *Venturia canescens* and its host, the moth *Plodia interpunctella*. The host here has five larval instars, and its generation length is around 41 days. The parasitoid lays eggs singly in the later instars of the host¹⁰; significant parasitoid development does not start until the host enters its final instar¹¹. The third species in the system is a pathogen, the *P. interpunctella* granulosis virus (PiGV), which infects larval hosts only, when consumed by them with their food. Lethal dose increases with larval instar, making instar V effectively invulnerable to infection¹². Data are reported here from replicated experimental laboratory systems comprising: (1) host and pathogen; (2) host and parasitoid; and (3) all three species together.

For host dynamics in host–pathogen systems (Fig. 1) the consistency between replicates was high. All populations fluctuated with marked regularity around an average density of roughly 150 moths, with a cycle period of 6 or 7 weeks, approximately one host generation (estimated by autocorrelation function (ACF) and spectral analysis of the time series). Fluctuations in the ACF, here and subsequently, were strongly damped with increasing

period, suggesting an endogenous cause for the fluctuations from within the populations¹³. The dynamics of infected larvae showed less clear evidence of regular periodicity (lower peaks of ACFs), although cycles comparable with those of the adult host are suggested in two replicates. Numbers of infected larvae, combined with a typical host fecundity of around 200 eggs¹⁴, suggest a prevalence of infection of around 1% and certainly less than 5%.

Host–parasitoid dynamics (Fig. 2) displayed regular fluctuations in both host and parasitoid, with periods similar to those in the host–pathogen systems. The parasitoid also significantly reduced host abundance (to a mean value of around 10), and substantially increased the amplitudes of the fluctuations.

Thus the host–pathogen and host–parasitoid systems both exhibited patterns of endogenous host generation-length cycles. A pattern of one-host-generation parasitoid–host cycles, as opposed to the classical multi-generation cycles, has been predicted by theory^{7,8} under the conditions pertaining in the *Plodia–Venturia* system¹⁵. The reduction of an order of magnitude in host density suggests strongly that these generation cycles were indeed host–parasitoid cycles. Host generation cycles have also been predicted for host–pathogen interactions when the host has an explicit stage structure⁹, as here. Previous empirical support for both these theoretical predictions has been scarce^{7,8,16,17}.

Host–pathogen–parasitoid populations were constructed by adding parasitoids to an existing host–pathogen interaction, either after many host–pathogen cycles (Fig. 1), or as soon as the host–pathogen system had become established (Fig. 3). The first notable contrast with the two-species patterns is the lack of persistence. All two-species populations persisted until they were either abandoned or used for other purposes, but of eight host–pathogen–parasitoid populations, seven failed to persist and the other was unavoidably terminated. In each extinction, the host disappeared first, followed by its predators. Natural populations would be affected by other features not considered here, such as spatial heterogeneity or further species, which may allow all three species to coexist.

Strikingly, the three-species system showed clear evidence of multi-generation cycles in the host and parasitoid. This is most apparent in Fig. 3a, where both species ran through four cycles of abundance, with significant peak ACFs at 20 weeks, before the extinction of the host. Three further replicates passed through two population cycles before host extinction (Fig. 3b, c) or termination (Fig. 3d), with similar peak periods for host (26, 25 and 21 weeks) and parasitoid (25, 25 and 17 weeks). The dynamics of virus-infected larvae in these populations showed no comparable cycles or regularities. A fifth replicate (Fig. 3e), and the three populations to which the parasitoid was added later (Fig. 1), went to extinction after a single rise and fall in parasitoid and host abundance, lasting a similar length of time to the cycles in the previous replicates (17–25 weeks for the parasitoid).

Overall, therefore, a remarkable picture emerges from the three-species system: cycles of abundance in both host and parasitoid, roughly 3–4 host generations in length, and of such amplitude that host extinction is likely following any one of them. Moreover, because the prevalence of overtly infected

* Present address: NERC Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 2SR, UK.

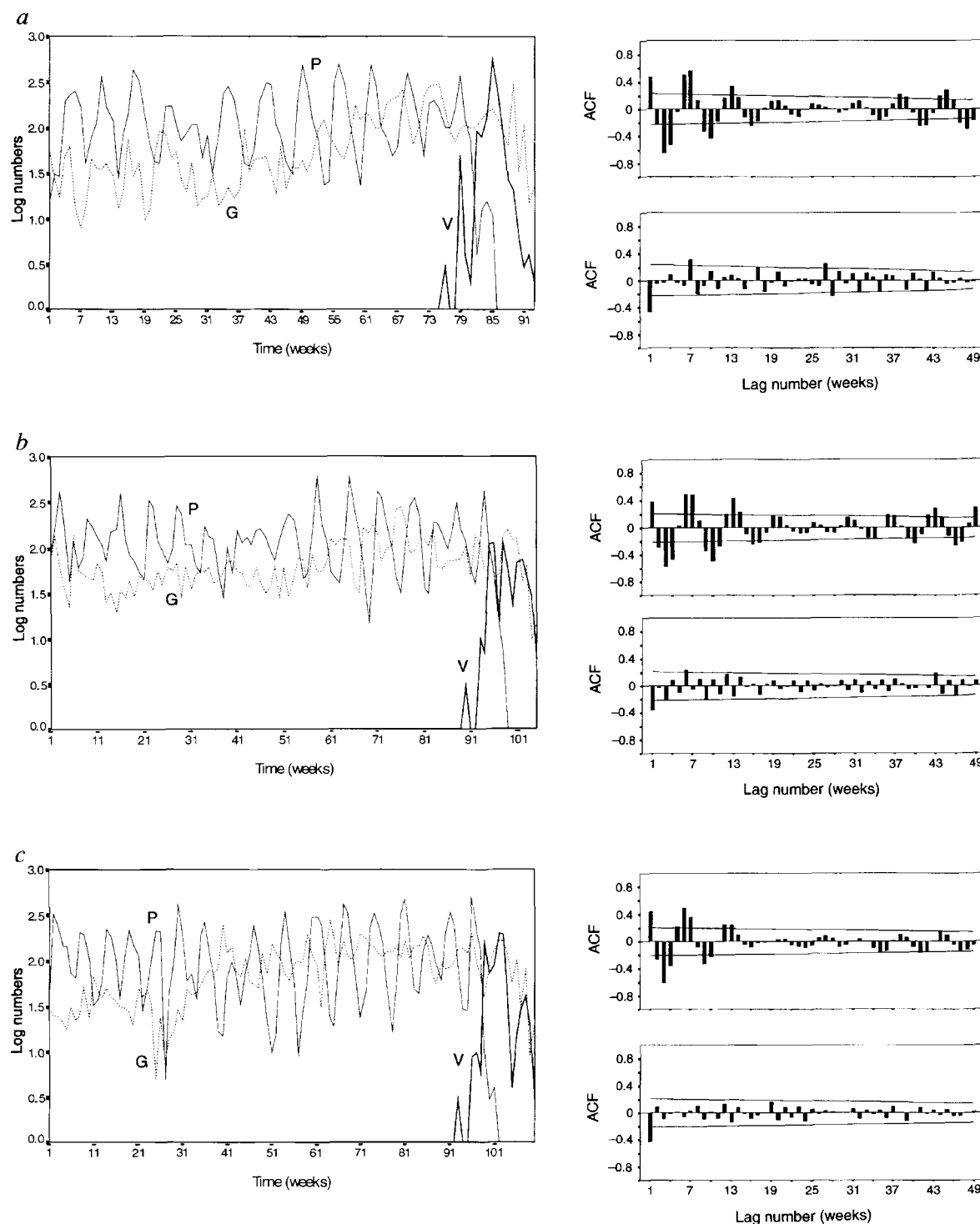


FIG. 1 Population dynamics (left) of adult *P. interpunctella* (thin line, P) and GV-infected larvae (broken line, G) (three replicates, following settling-down periods of 301, 217 and 182 days, respectively). Autocorrelation functions (ACFs, right) for the host (top) and the infected larvae (bottom) were calculated to establish the period, strength and consistency of population cycles. Sloping lines show the values the bars must exceed for statistical significance ($P < 0.05$). Cyclicity is evidenced by repeated patterns in ACFs (not single significant values), and significant values at or near lag-1 are ignored (suggesting only that consecutive values are correlated)²⁸. Spectral analysis was also performed to determine the cycle period more precisely. Analyses were carried out on logged data and, for the infected larvae, to de-trend a significant increase, on differences between successive (logged) data points rather than on the data points themselves. At the end of each data set are the dynamics of the adult parasitoid, *V. canescens* (bold line, V) after it had been allowed to invade the population

cages (two newly emerged parthenogenetic females) 800 days after initial establishment. Before the introduction of the parasitoid, mean host densities (untransformed data $\pm$ s.e.) were: a, 153.5 ± 14.4 ; b, 151.1 ± 12.6 ; c, 134.7 ± 11.6 . The cycle periods (from ACF and spectral analysis) and the ACFs ($P < 0.05$) of those periods for the adult host (H) and infected larval (GV) populations were: a, H, 7, 6.7, 0.58; GV, 7, 6.2, 0.32; b, H, 6, 6.8, 0.51; GV, 6, 6.1, 0.24; c, H, 6, 6.4, 0.52; GV, no significant peak. Each population was initiated with 15 male and 15 female fifth-instar larvae, plus twelve heavily infected fifth-instar larvae, in closed perspex cages ($22 \times 22 \times 9$ cm). In an established protocol^{15,16,29} a one-sixth section of the medium was replaced sequentially with fresh food each week, at which point all dead adults in the population were removed and counted. The dynamics of PiGV infection were monitored by counting the numbers of infected larvae in the discarded food.

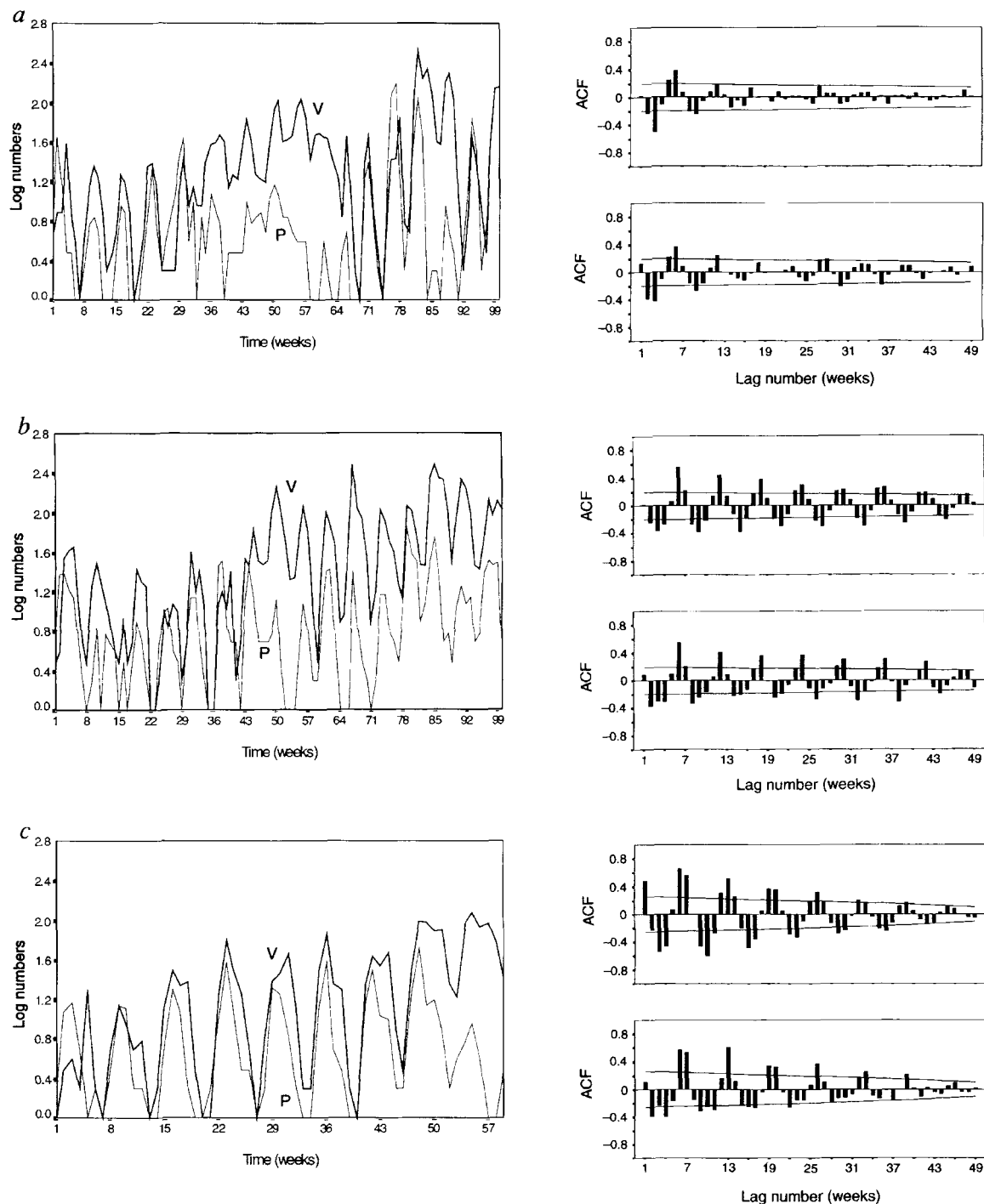


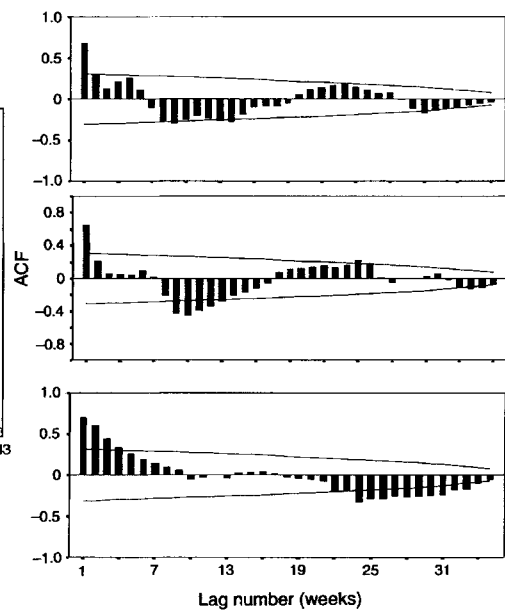
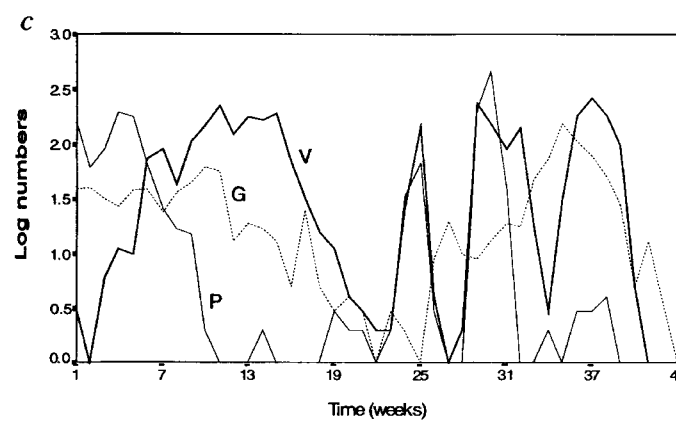
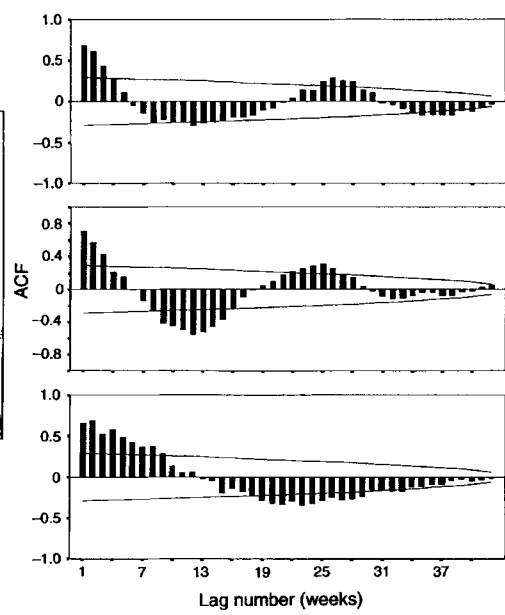
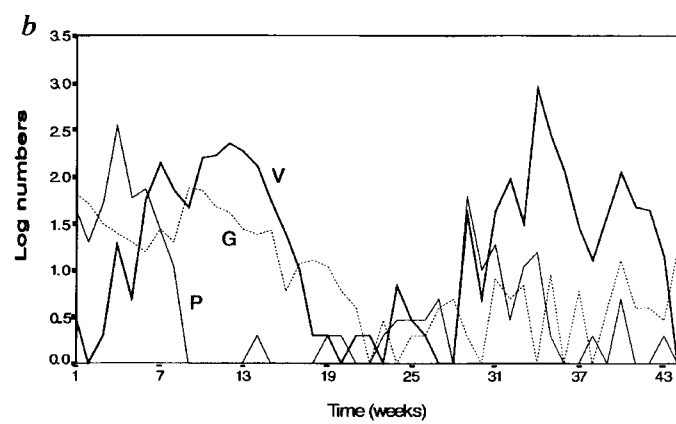
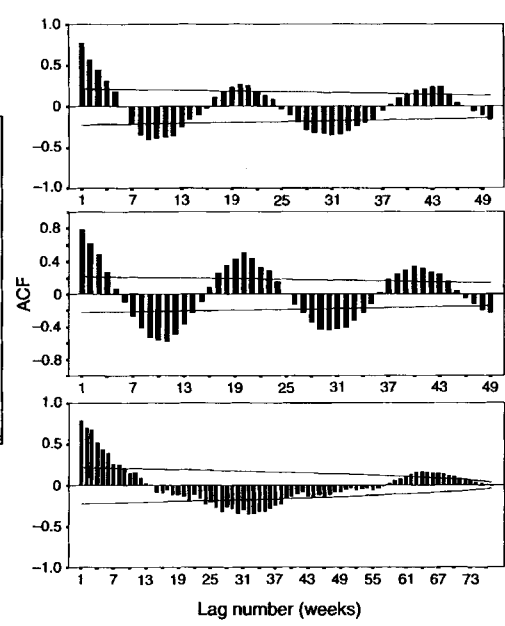
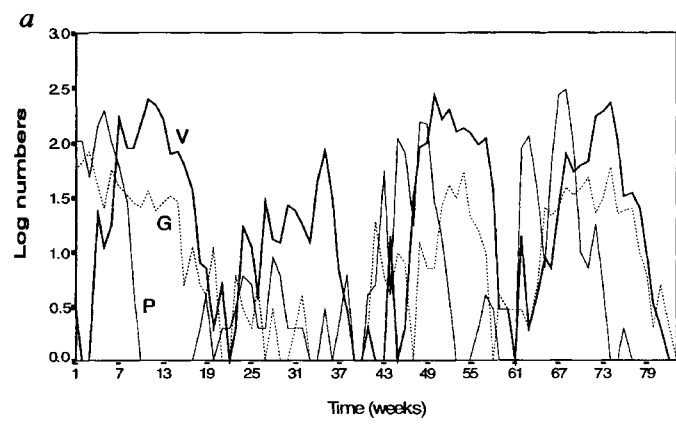
FIG. 2 Population dynamics of *P. interpunctella* (thin line P) and *V. canescens* (bold line, V) (three replicates) with their corresponding ACFs for the host (top) and parasitoid (bottom) (performed on logged data). Methods were as described in Fig. 1, except that two newly emerged parasitoids (parthenogenetic females) were added to the cages 14 days after the initial host population had mated, and dead adults of both the host and parasitoid were counted each week. For each adult host (H) and

parasitoid (P) population, mean population densities ($\pm$ s.e.) following a settling down period (brackets) were: a, H, 14.0 ± 3.1 ; P, 38.5 ± 5.5 (56 days); b, H, 9.6 ± 1.2 ; P, 55.4 ± 6.5 (49 days); c, H, 7.6 ± 1.4 ; P, 27.1 ± 4.1 (49 days). The cycle periods (from ACF and spectral analysis) and the ACFs ($P < 0.05$) of those periods were: a, H, 6, 5.8, 0.39; P, 6, 6.0, 0.39; b, H, 6, 6.3, 0.53; P, 6, 6.5, 0.56; c, H, 6, 6.5, 0.66; P, 6, 6.4, 0.55.

hosts was low (Fig. 1), and host-alone dynamics are only weakly affected by the presence or absence of infection¹⁶, it is likely that, if these dynamics had been observed in natural populations, they would have been described as those of a simple host–parasitoid interaction, rather than host–parasitoid dynamics within a three-species system.

A mechanistic model of host–pathogen–parasitoid interactions¹⁸ has predicted multi-generation cycles (amongst other

patterns), but the model assumes a highly seasonal environment not applicable to the present system, and shows no tendency for generation cycles in the component interactions (host–pathogen and host–parasitoid). Models of ‘intraguild predation’¹⁹ (prey, plus predator, plus second predator that consumes both first predator and prey) also predict multi-generation cycles²⁰, but again there are no generation cycles in component systems, and the assumed asymmetry between predators is inappropriate



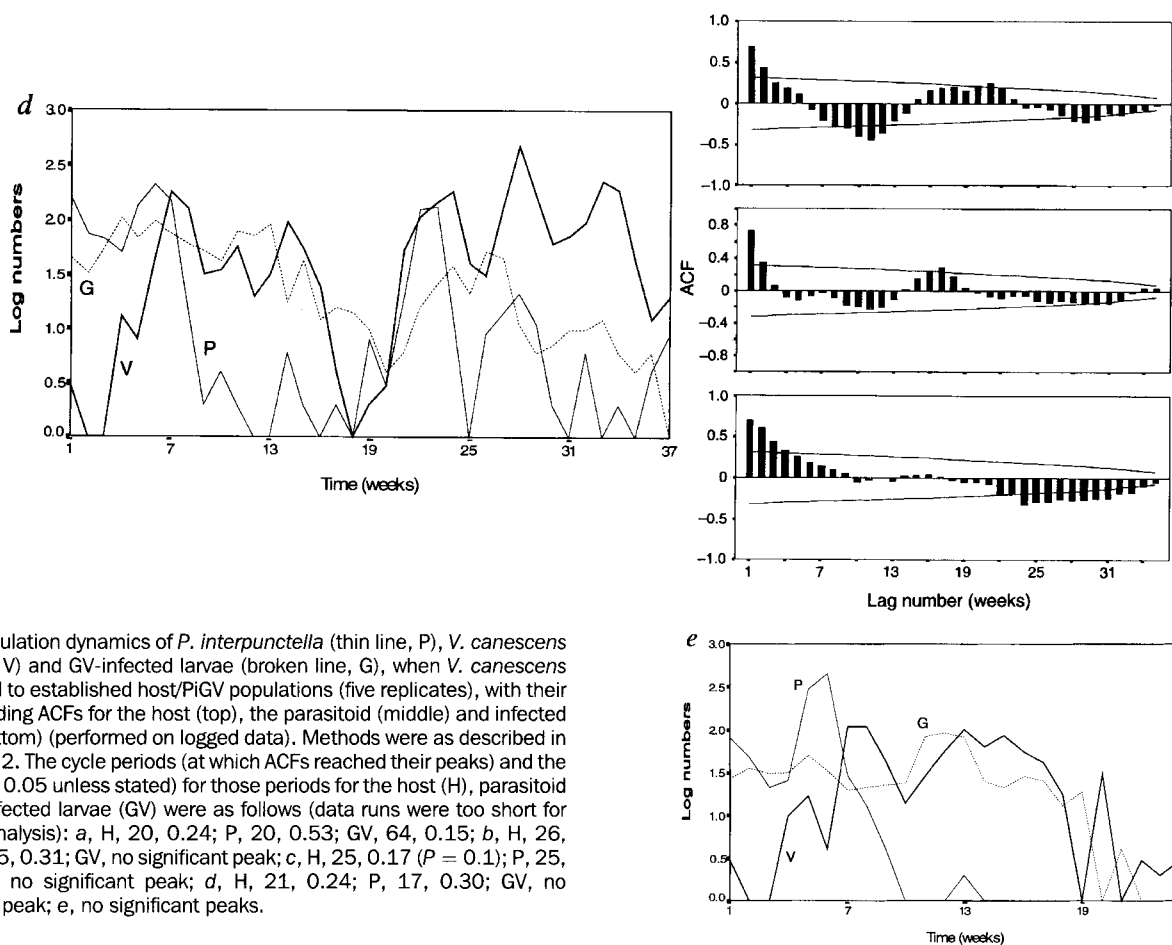


FIG. 3 Population dynamics of *P. interpunctella* (thin line, P), *V. canescens* (bold line, V) and GV-infected larvae (broken line, G), when *V. canescens* was added to established host/PIGV populations (five replicates), with their corresponding ACFs for the host (top), the parasitoid (middle) and infected larvae (bottom) (performed on logged data). Methods were as described in Figs 1 and 2. The cycle periods (at which ACFs reached their peaks) and the ACFs ($P < 0.05$ unless stated) for those periods for the host (H), parasitoid (P) and infected larvae (GV) were as follows (data runs were too short for spectral analysis): a, H, 20, 0.24; P, 20, 0.53; GV, 64, 0.15; b, H, 26, 0.27; P, 25, 0.31; GV, no significant peak; c, H, 25, 0.17 ($P = 0.1$); P, 25, 0.24; GV, no significant peak; d, H, 21, 0.24; P, 17, 0.30; GV, no significant peak; e, no significant peaks.

here²¹. A more plausible explanation, which remains to be fully quantified, is that generation cycles in the component systems arise, as predicted by theory^{7,9}, from the observed concentrations of mortality in particular stages of the host's life cycle (pathogen, earlier larval instars; parasitoid, later instars^{10,12,21}). However, when the host interacts with a 'conglomerate' predator, comprising both pathogen and parasitoid, they exert, between them, no markedly stage-structured effects, giving rise to multi-generation cycles, as predicted by classical models (lacking stage structure).

Apparent empirical demonstrations of multi-generation predator-prey cycles have been rare³⁻⁶. This might imply that they are only observed when the intensity of interaction is sufficient to generate cycles, despite there being a web of species interactions to muddy the waters. The present results, however, raise the possibility that this may be exactly contrary to the truth, because we found that multi-generation cycles in a predator

(the parasitoid) and its prey are apparent only when there are further interactions with the pathogen. Alone, host and parasitoid exhibit even clearer cycles, but these are roughly one and not several host generations long. From this alternative perspective, the predator-prey interaction, far from being of the essence in generating multi-generation 'predator-prey' cycles, is actually irrelevant in its own right, assuming importance only as an integral part of a more complex interaction. Our data therefore support the view that studies of simple multi-species systems can be both feasible and capable of generating novel dynamical patterns²²⁻²⁶.

The case that population ecology has much to gain by studying transitions in dynamics induced by changes in demographic parameters has recently been well made²⁷. Clearly, studying transitions induced by changes in the structure of simple, tractable systems may be equally challenging and productive. □

Received 21 November 1995; accepted 21 March 1996.

1. Lotka, A. J. *Elements of Physical Biology* (Williams and Wilkins, New York, 1925).
2. Volterra, V. J. *Cons. perm. Int. Explor. Mer.* **3**, 3-51 (1928).
3. Begon, M., Harper, J. L. & Townsend, C. R. *Ecology: Individuals, Populations and Communities* 3rd edn (Blackwell Science, Oxford, 1996).
4. Myers, J. H. *Adv. Ecol. Res.* **18**, 179-242 (1988).
5. Hanski, I., Turchin, P., Korpimäki, E. & Henttonen, H. *Nature* **364**, 232-235 (1993).
6. Krebs, C. J. et al. *Science* **269**, 1112-1115 (1995).
7. Godfray, H. C. J. & Hassell, M. P. *J. Anim. Ecol.* **58**, 153-174 (1989).
8. Gordon, D. M., Nisbet, R. M., De Roos, A., Gurney, W. S. C. & Stewart, R. R. *J. Anim. Ecol.* **60**, 295-308 (1991).
9. Briggs, C. J. & Godfray, H. C. J. *Am. Nat.* **145**, 855-887 (1995).
10. Sait, S. M. et al. *Ecol. Ent.* **20**, 199-201 (1995).
11. Harvey, J. A., Harvey, I. F. & Thompson, D. J. *Ecology* **75**, 1420-1428 (1994).
12. Sait, S. M., Begon, M. & Thompson, D. J. *J. Invert. Pathol.* **63**, 107-110 (1994).
13. Turchin, P. & Taylor, A. D. *Ecology* **73**, 289-305 (1992).
14. Sait, S. M., Begon, M. & Thompson, D. J. *J. Anim. Ecol.* **63**, 541-550 (1994).
15. Begon, M., Sait, S. M. & Thompson, D. J. *Proc. R. Soc. Lond. B* **260**, 131-137 (1995).
16. Sait, S. M., Begon, M. & Thompson, D. J. *J. Anim. Ecol.* **63**, 861-870 (1994).
17. Reeve, J. D., Cronin, J. T. & Strong, D. R. *J. Anim. Ecol.* **63**, 912-920 (1994).

18. Hochberg, M. E., Hassell, M. P. & May, R. M. *Am. Nat.* **135**, 74-94 (1990).
19. Polis, G. A. & Holt, R. D. *Trends Ecol. Evol.* **7**, 151-154 (1992).
20. Holt, R. D. in *Multitrophic Interactions* (eds Gange, A. C., Brown, V. K. & Jones, T. H.) (Blackwell Science, Oxford, 1996) (in the press).
21. Begon, M., Bowers, R. G., Sait, S. M. & Thompson, D. J. in *Frontiers of Population Ecology* (eds Floyd, R. B. & Sheppard, A. W.) (CSIRO Press, Melbourne, 1996) (in the press).
22. Vandermeer, J. H. *Ecology* **50**, 362-371 (1969).
23. Wilbur, H. M. & Alford, R. A. *Ecology* **66**, 1106-1114 (1985).
24. Muratori, S. & Renaldi, S. *SIAM J. appl. Math.* **52**, 1688-1706 (1992).
25. Lawler, S. P. & Morin, P. J. *Am. Nat.* **141**, 675-686 (1993).
26. Morin, P. J. & Lawler, S. P. *A. Rev. Ecol. Syst.* **26**, 505-529 (1995).
27. Costantino, R. F., Cushing, J. M., Dennis, B. & Desharnais, R. A. *Nature* **375**, 227-230 (1995).
28. SPSS for Windows: *Trends, Release 6.0* (SPSS, Chicago, 1993).
29. Gurney, W. S. C., Nisbet, R. M. & Lawton, J. H. J. *J. Anim. Ecol.* **52**, 479-495 (1983).

ACKNOWLEDGEMENTS. We thank C. Godfray and I. Hanski for comments on the manuscript. This work was supported by the NERC.

CORRESPONDENCE and requests for materials should be addressed to M.B. (e-mail: mbegon@liv.ac.uk).

Specification of the wing by localized expression of *wingless* protein

Medard Ng, Fernando J. Diaz-Benjumea, Jean-Paul Vincent*, Jun Wu† & Stephen M. Cohen

European Molecular Biology Laboratory, Postfach 10.2209, 69012 Heidelberg, Germany

* Laboratory for Molecular Biology, Medical Research Council, Cambridge CB2 2QH, UK

† Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA

Limb development in *Drosophila* depends on subdivision of the limb primordia into functional units called compartments¹⁻⁴. Cell interactions across compartment boundaries establish pattern-organizing centres that control growth and specify cell fates along the anteroposterior (AP) and dorsoventral (DV) axes of the limbs^{2,3,5-9}. AP subdivision of the disc primordia is inherited from the embryonic ectoderm¹⁰. DV subdivision of the wing disc occurs during the second larval instar through localized expression of the *apterous* protein (*Apterous*) in dorsal cells^{2,11}. A third major subdivision of the wing disc into wing and body-wall compartments also occurs in the second instar¹. Here we show that specification of the wing primordium in early second instar depends on activity of the AP patterning system but not the DV system. These results define two distinct roles for the *wingless* gene: a primary role in specifying the wing primordium, and a subsequent role mediating the patterning activities of the DV compartment boundary^{9,12,13}.

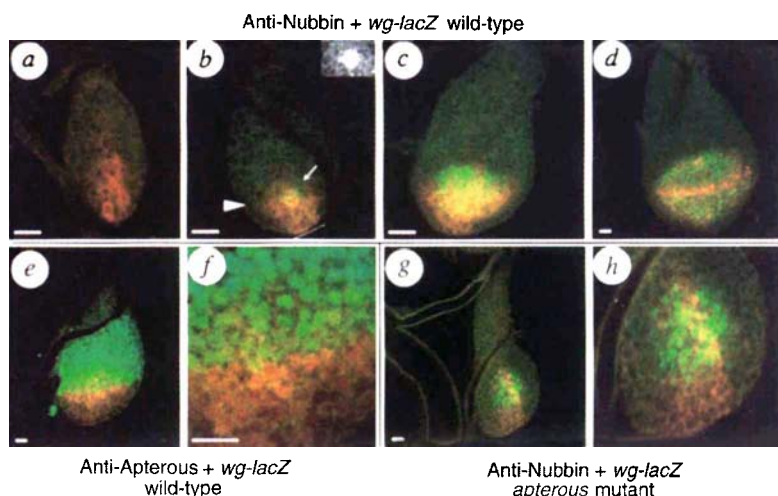
The wing imaginal disc comprises the primordia of both adult wing and body wall. Studies of compartment formation¹ and *wingless* (*wg*) function^{14,15} suggest that the developmental decision between wing and body wall is made in second larval instar. To investigate how the wing primordium is specified as a distinct population of cells within the imaginal disc, we sought wing-specific markers expressed at this stage. Although *vestigial* activity is required for wing formation, expression of *vestigial* protein is not

restricted to cells of the wing primordium, either in the embryonic wing disc, or later when the *vestigial* protein domain resolves to a stripe at the DV compartment boundary^{6,16,17}. The gene *nubbin* encodes a POU-domain protein also required for wing formation¹⁸. In contrast to *Vestigial*, expression of *Nubbin* is first detected in second instar in a group of cells that appears to correspond to the newly specified wing primordium (Fig. 1).

To examine early stages in the formation of the wing primordium, we compared expression of *wg* and *nubbin* proteins (Fig. 1). We found that *wg* is expressed in a dynamic pattern that reflects the AP and DV subdivision of the disc^{15,16}, and is first expressed in a wedge of cells in the anterior compartment. *Nubbin* is not detectable at this stage (Fig. 1a). Shortly thereafter, a new domain of *wg* expression appears as a band one or two cells in width (arrows in Fig. 1b). This *wg* domain crosses the AP boundary and includes cells in both compartments. It consists of ventral cells adjacent to the DV boundary (Fig. 1e, f). In these cells, *wg* expression depends on *apterous* activity (Fig. 1g, h). *Nubbin* is first expressed in a small group of cells located where the second *wg* protein domain overlaps the AP boundary (Fig. 1b). Expression of *wg* undergoes several modulations during second instar^{15,16}, during which *wg* and *nubbin* are expressed coextensively throughout the wing pouch (Fig. 1c, d).

As *Nubbin* is expressed near the interface between AP and DV compartment boundaries, we investigated whether the activities of these patterning systems are required to initiate its expression. There was no expression of *Nubbin* in wing discs from which *wg* or *hedgehog* activity is removed in mid-first instar (Fig. 2a, b); the requirement for *hedgehog* is mediated through *wg*, (Fig. 2a legend). Although formation of the wing primordium depends on *wg*, it does not require activity of the DV system. The wing is present as a discrete population of *Nubbin*-expressing cells in discs lacking *apterous* or *vestigial* function (Fig. 2c, d). The secondary modulations of the *wg* pattern described above are not observed in discs mutant for *apterous* (Fig. 1, compare c, d with g, h). Nonetheless, *Nubbin* is expressed in *apterous* mutant discs, in a domain centred on the primary wedge of *wg* expression (Fig. 1g, h). These results indicate that specification of the wing depends on the primary domain of *wg* expression in the anterior compartment, and not on the secondary domains of *wg*, which themselves depend on the DV system. We note that the ability of *wg* protein (*Wg*) to activate *nubbin* is spatially restricted; cells at the periphery of the

FIG. 1 Expression of *Nubbin* and *Wingless* define the wing primordium (a–d, g, h); labelling: green for *Nubbin* (a–d, g, h) and *Apterous* (e, f) proteins; red, *wg-lacZ*. Wild-type (a–f) and *apterous* mutant (g, h). Anterior, left; dorsal, top. Relative magnification is indicated by scale bars. a, Early second instar wing disc; *wg* is expressed in an anterior wedge, *nubbin* is not expressed. b, Slightly older second instar disc; *Nubbin* is first detected as a small cluster of cells where the two domains of *wg* expression intersect (small arrow, *nubbin*; inset, *nubbin* alone; arrowhead, second *wg* domain; thin bracket, *wg* anterior wedge). c, Mid second instar; *nubbin* expression fills the wing pouch; *wg* is expressed more strongly in the ventral pouch. (Dorsal *wg* expression is partly obscured by *Nubbin*, but is visible in the single label; not shown). In the anterior compartment the DV distinction is obscured owing to perdurance of β -gal from the anterior wedge. At this stage, expression of *vestigial* protein fills the entire wing pouch and extends into the notum¹⁶. d, Late second instar; the ventral *wg* domain resolves into a stripe at the DV boundary. e, f, Mid second instar; *apterous* protein marks the dorsal compartment². By comparison, the elevated level of *wg* expression is restricted to the ventral compartment. e, Intermediate in age between c and d. The anterior wedge of *wg* extends well into the dorsal compartment (compare c and e) and so cannot be considered ventral. g, h, Expression of *Nubbin* in a late second instar *apterous* mutant disc. The primary wedge of *wg-lacZ* persists, but the secondary domains do not form; *Nubbin* is expressed in anterior and posterior compartments. The level of *nubbin* expression is



lower in *apterous* mutant discs, suggesting that secondary *wg* expression may help to increase the level of *nubbin* expression. Magnification of h is twice that of g.

METHODS. Antibody was raised against an N-terminal fragment of *nubbin* protein in rats. Anti-*apterous* was raised as described in ref. 22.

disc do not turn on *nubbin* despite expression of *wg* protein. This suggests that a second activity limits the competence of cells to respond to *wg* protein by activating *nubbin*, and that their combined activity defines the size and position of the wing field.

The proposal that *wg* specifies the wing primordium is substantiated by the finding that ectopic expression of Wg can induce supernumerary wings in the portion of the disc normally fated to give rise to body wall (Fig. 3). The ability of Wg protein to

reprogram cells in the notum to wing pouch identity is limited to a very early stage in wing development (Fig. 3b legend). Based on the dynamics of *wg* expression, we infer that requirement for *wg* activity in specifying the wing pouch is most likely to be restricted to the early second instar, when *nubbin* is first expressed. The supernumerary wing expresses Wg in a stripe along the presumptive wing margin (Fig. 3a–c) and is bisected by *apterous* expression (Fig. 3d–f). Note that the level of *wg* protein expressed in the

FIG. 2 Expression of Nubbin in mutant wing discs. **a**, Third-instar discs mutant for *hedgehog* (*hh*); *hh^{ts2}/hh^{U35}* larvae were raised at 18 °C and shifted to 30 °C to inactivate *hedgehog* protein in mid first instar (36 ± 12 h). The absence of Nubbin expression indicates that the wing is missing. Overstaining to bring out the weaker Nubbin expression in the leg resulted in nonspecific background, but this is not nuclear (compare with **c**, **d**). There is no expression of *wg* in *hh^{ts2}* homozygous flies or in *hh^{ts2}/hh^{U35}* discs shifted to 30 °C in mid first instar (36 ± 12 h; data not shown). Removing *hh* activity in mid second instar³ (60 h) does not eliminate *wg* expression, and does not lead to loss of the wing pouch. **b**, Third-instar discs mutant for *wingless*; *wg^{ts}/wg^{ts3}* larvae were raised at 18 °C and shifted to 30 °C to inactivate Wg in mid first instar (36 ± 12 h). The absence of Nubbin expression indicates that the wing is missing. Staining discs of this genotype with antibody to *vestigial* protein shows that the wing is replaced by a duplication of the notum (asterisk; see Fig. 4f). Because the combined action of Wg and Decapentaplegic protein is required to define the leg primordium by activating *Distal-less* expression⁵, we investigated whether *dpp* activity is required for *nubbin* protein expression. Combinations of *dpp* mutant alleles that eliminate *Distal-less* expression in the leg do not eliminate Nubbin in the wing (not shown). **c**, **d**, Expression of Nubbin shows the presence of small wing pouches in third-instar discs homozygous for the null allele *apterous*^{UG035} (**c**) or *vestigial*^{B3627} (**d**). Magnification of **a** is twice that of **b–d**.

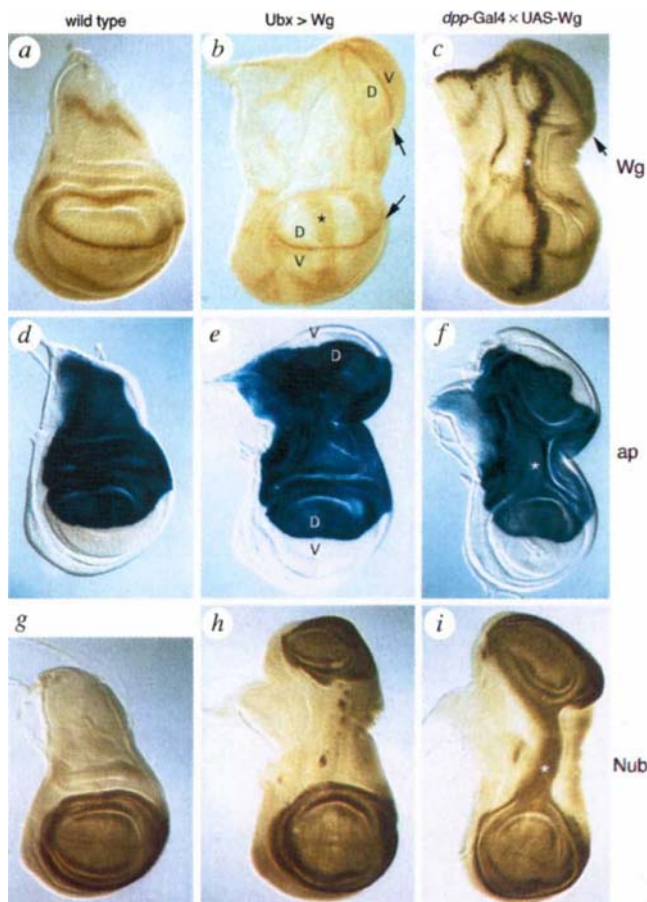
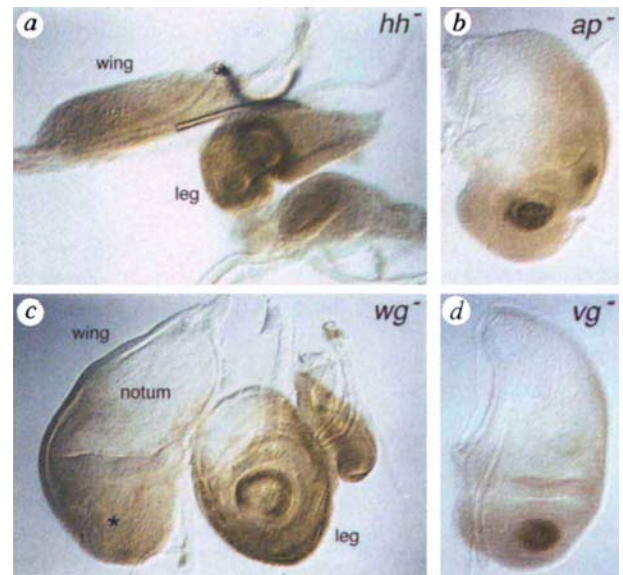


FIG. 3 Supernumerary wings were induced by misexpression of *wg* protein in a defined spatial pattern using the GAL4 system (**c**, **f**, **i**) or by producing clones of *wg*-expressing cells (**b**, **e**, **h**). All discs are third instar. D, dorsal wing pouch; V, ventral wing pouch. **a–c**, Expression of Wg in the wing margin of endogenous and supernumerary wings. **a**, Supernumerary wings induced by clones of *wg*-expressing cells; *wg* expression in the clones (one is indicated by an asterisk) is lower than at the wing margin (arrows). The endogenous *wg* gene has been turned on in the supernumerary wing margin. Extra wings are formed when *wg*-expressing clones are induced at 36 ± 6 h but not at 48 ± 6 h. Discs showing duplicated wings are recovered at high frequency, but few larvae survive to adulthood. There is no consistent pattern to the position or amount of *wg*-expressing tissue in adult supernumerary wings (not shown). **c**, Supernumerary wing induced by *wg* expression under control of *dpp*-GAL4 (asterisk). The endogenous *wg* gene is expressed in the supernumerary wing margin (arrow). Adults with wing duplications can be recovered when larvae are raised at 18 °C to reduce GAL4 activity (not shown). **d–f**, Expression of *apterous* subdivides endogenous and supernumerary wing pouches into dorsal and ventral compartments. **f**, The supernumerary wing arises where the Wg stripe crosses the *apterous* domain in the notum, hence the reversal of DV polarity with respect to the endogenous wing. Supernumerary wing tissue (produced by *wg*-expressing clones) can also form at more lateral positions. In such cases the supernumerary and endogenous wings usually fuse (not shown). **g–i**, Expression of Nubbin in endogenous and supernumerary wing pouches. **h**, Note small patches of Nubbin expression induced by clones in the notum in addition to the supernumerary wing. **i**, The endogenous and supernumerary wings are connected by a band of *nubbin*-expressing cells straddling the domain of ectopic *wg* protein expression. Examining adult wings shows that this represents an expansion of the wing hinge to connect the two wing pouches (not shown).

METHODS. *Ubx > f⁺ > Wg* and anti-Wg, ref. 9; *dpp*-GAL4, ref. 23; UAS-Wg, ref. 24. To induce clones of Wg-expressing cells, larvae of genotype *yHsflp; Ubx > f⁺ > Wg/+* were subjected to 30 min heat-shock treatment at 38 °C at 36 ± 6 h (mid first instar) or at 48 ± 6 h (early second instar).

supernumerary wing margin is higher than its level in the clones of *wg*-expressing cells (Fig. 3b), but lower than its level under control of the GAL4 system (Fig. 3c). These observations suggest that interaction between *apterous*-expressing dorsal cells and ventral cells in the ectopic wing pouch leads to activation of endogenous *Wg* at the DV boundary of the supernumerary wing.

We propose that specification of the wing pouch is required to define a field of cells in which the DV patterning system can function to control the growth and patterning of the wing. Well-formed supernumerary wings are always positioned over the DV compartment boundary. Note, however, that ectopic expression of *Wg* can also respecify cell fate internally in the notum, as indicated by the continuous band of cells expressing Nubbin along the *Wg* stripe in Fig. 3i. With the decapentaplegic-GAL4 driver, the ectopic expression of *Wg* meets the edge of the dorsal compartment on the opposite side of the notum from the endogenous wing (Fig. 3c). Consequently, the DV polarity of the supernumerary wing is opposite to that of the endogenous wing (Fig. 3f). Super-

numerary wings produced by clones of *wg*-expressing cells in a comparable position in the disc show a similar relation to the endogenous wing.

Files homozygous for *wg*¹, a viable mutation, often lack wings and show a symmetric duplication of body-wall structures^{14,19}. Reducing *wg* activity in the second instar produces a similar phenotype¹⁵. Clonal analysis suggested that the duplicated body wall arises by transformation of wing progenitor cells to notum identity¹⁴. In view of our finding that early misexpression of *wg* can reprogram notum to wing identity, the *wg*¹ phenotype could be explained by failure to specify the wing pouch in the early second-instar disc. Consistent with this prediction, Nubbin is not expressed in second instar wing discs from *wg*¹ mutants (Fig. 4c, d). Instead, the region where the pouch should have been specified expresses *teashirt* protein, which is normally restricted to the notum (Fig. 4a, b). Failure to specify the wing pouch in the second instar leads to a symmetric duplication of body-wall structures in the mature disc (visualized by expression of *vestigial* protein; Fig. 4e, f; refs 14,16). Conversely, ectopic expression of *Wg* in early second instar leads to symmetric duplication of the wing (Fig. 3). These findings suggest that the early function of *Wg* is to specify the wing primordium, and that failure to do so causes cells to adopt a default identity of body wall.

The results described here allow us to define two distinct functions for *wg* protein in wing formation: a primary role in specification of the wing primordium, and a secondary role as a mediator of the growth and patterning activities of the DV compartment boundary⁹. In the absence of the DV system, the wing primordium is specified but remains small in size²⁰ (Figs 1 and 2). This contrasts with *wg* mutants in which no wing is specified^{14,15} (Figs 2 and 4). Specification of the wing is accomplished without any secondary modulation of the *wg* pattern. The fact that even the earliest secondary modulations of *wg* expression are *apterous*-dependent is consistent with the observation that *wg* expression at the DV boundary depends on a signal from dorsal to ventral cells mediated by *Serrate* and *Notch* proteins⁹. We conclude that ventral expression of *Wg* is a consequence of activity of the DV system, and therefore downstream of the *Serrate* and *Notch* proteins. Our findings contrast with the proposal that the ventrally expressed *wg* and dorsally expressed *Serrate* proteins work together to specify the wing through activation of *Vestigial* at the DV boundary²¹. Rather, we view both *wg* and *Vestigial* proteins as targets for activation by the DV system; both are required to promote the growth of the wing, but only after the wing field has been established. □

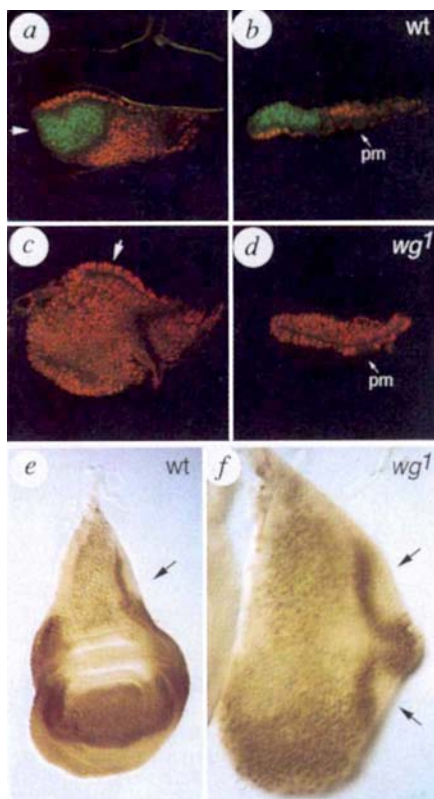


FIG. 4 Failure to specify the wing pouch causes transformation to notum cell identity. *a, b*, Wild-type second instar wing disc labelled for Nubbin (green, expressed in the wing pouch) and *teashirt* protein (red, expressed in the notum). *a*, Horizontal optical section, and *b*, optical section perpendicular to the plane of the disc shown in *a* in the position indicated by the white arrow; pm indicates *teashirt* expression in the peripodial membrane. *c, d*, *wg*¹ mutant second instar wing disc labelled for *nubbin* (green) and *teashirt* (red) proteins. Note the absence of Nubbin protein and the expanded distribution of *Teashirt*. Arrow indicates the plane of the perpendicular optical section in *d*. Anterior, top; dorsal, right, in *a–d*. *e, f*, Expression of *vestigial* protein in wild-type and *wg*¹ mutant discs in third instar. The large labelled cells in the notum are ad epithelial cells located under the plane of the disc epithelium^{16,25}. The arrow indicates the strip of *Vestigial* expression along the DV boundary in the notum. *f*, Note the symmetric duplication of the notum stripe of *vestigial* protein (arrows), the absence of the wing pouch, and the expanded field of ad epithelial cells in the duplicated notum of the *wg*¹ mutant disc. Anterior, left; dorsal, top, in *e* and *f*. Magnification of *f* is twice that of *e*.

METHODS. Antibody to full-length *teashirt* protein was raised in mice.

Received 15 November 1995; accepted 21 March 1996.

- García-Bellido, A., Ripoll, P. & Morata, G. *Nature new Biol.* **245**, 251–253 (1973).
- Diaz-Benjumea, F. J. & Cohen, S. M. *Cell* **75**, 741–752 (1993).
- Basler, K. & Struhl, G. *Nature* **368**, 208–214 (1994).
- Blair, S. S. *Bioessays* **17**, 299–309 (1995).
- Diaz-Benjumea, F. J., Cohen, B. & Cohen, S. M. *Nature* **372**, 175–179 (1994).
- Williams, J. A., Paddock, S. W., Vorwerk, K. & Carroll, S. B. *Nature* **368**, 299–305 (1994).
- Tabata, T. & Kornberg, T. *Cell* **76**, 89–102 (1994).
- Zecca, M., Basler, K. & Struhl, G. *Development* **121**, 2265–2278 (1995).
- Diaz-Benjumea, F. J. & Cohen, S. M. *Development* **121**, 4215–4225 (1995).
- Cohen, S. M. *Nature* **343**, 173–177 (1990).
- Blair, S. S. *Development* **119**, 339–351 (1993).
- Phillips, R. & Whittle, J. R. S. *Development* **118**, 427–438 (1993).
- Couso, J. P., Bishop, S. & Martinez Arias, A. *Development* **120**, 621–636 (1994).
- Morata, G. & Lawrence, P. A. *Dev Biol.* **56**, 227–240 (1977).
- Couso, J. P., Bate, M. & Martinez Arias, A. *Science* **259**, 484–489 (1993).
- Williams, J. A., Paddock, S. W. & Carroll, S. B. *Development* **117**, 571–584 (1993).
- Cohen, B., Simcox, A. A. & Cohen, S. M. *Development* **117**, 597–608 (1993).
- Ng, M., Diaz-Benjumea, F. J. & Cohen, S. M. *Development* **121**, 589–599 (1995).
- Sharma, R. P. & Chopra, V. L. *Dev Biol.* **48**, 461–465 (1976).
- Speicher, S. A., Thomas, U., Hinz, U. & Knust, E. *Development* **120**, 535–544 (1994).
- Couso, J. P., Knust, E. & Martinez Arias, A. *Curr. Biol.* **5**, 1437–1448 (1995).
- Kim, J., Irvine, K. D. & Carroll, S. B. *Cell* **82**, 795–802 (1995).
- Wilder, E. L. & Perrimon, N. *Development* **121**, 477–488 (1995).
- Lawrence, P. A., Bodmer, R. & Vincent, J. P. *Development* **121**, 4303–4308 (1995).
- Williams, J. A., Bell, J. B. & Carroll, S. B. *Genes Dev.* **5**, 2481–2495 (1991).

ACKNOWLEDGEMENTS. We thank S. Eaton, W. Brook and C. Neumann for suggesting improvements to the manuscript; J. Kim and S. Carroll for *Apterous* and *Vestigial* antibody; and E. Wilder for *dpp-Gal4*.

CORRESPONDENCE and requests for materials should be addressed to S.M.C. (e-mail: scohen@embel-heidelberg.de).

Function of the Eph-related kinase *rtk1* in patterning of the zebrafish forebrain

Qiling Xu*, Graham Alldus†, Rachel Macdonald*, David G. Wilkinson† & Nigel Holder*

* DBRC, The Randall Institute, King's College, 26–29 Drury Lane, London WC2B 5RL, UK

† Division of Developmental Neurobiology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

EARLY during its development, the vertebrate brain is subdivided into regions that have distinct fates and correlate with the expression domains of regulatory genes^{1,2}, but little is known about the cell–cell interactions that establish this spatial pattern. Candidates for regulating such interactions are the Eph-related receptor tyrosine kinases (RTKs) which have spatially restricted expression in the developing brain^{2–6}. These RTKs may mediate cell-contact-dependent signalling by interacting with membrane-bound ligands⁷, and have been implicated in axon repulsion^{8,9} and the segmental restriction of gene expression in the hindbrain¹⁰, but nothing is known regarding their function in the rostral neural epithelium. Here we use a dominant-negative approach in the zebrafish embryo to interfere with the function of *Rtk1*, an Eph-related RTK expressed in the developing diencephalon. We find that expression of a truncated receptor leads to expansion of the eye field into diencephalic territory and loss of diencephalic structures, indicating a role for *Rtk1* in patterning the developing forebrain.

Like its murine counterpart *Sek-1* (refs 6,11), zebrafish *rtk1* is expressed in specific regions of the developing mesoderm, fore-

brain and hindbrain^{2,10,12}. At the end of gastrulation *rtk1* is expressed in the lateral part of the presumptive forebrain and these expression domains then converge towards the midline as the neural keel forms (Fig. 1a, c). Transverse sections reveal that expression of *rtk1* is restricted to the presumptive diencephalon and does not occur in the optic primordium (Fig. 1b, d). Subsequently, the expression in the diencephalon is restricted to the presumptive dorsal thalamus and the region adjacent to the optic recess of the third ventricle (Fig. 1e, f).

To test the function of *rtk1*, we undertook a dominant negative approach analogous to that used for other receptor kinases^{13–15}. In previous studies we found that microinjection of RNA encoding a truncated *Sek-1* receptor into cleavage-stage zebrafish embryos disrupts patterning of hindbrain segments¹⁰. Analysis of embryos at 5 days of development revealed that there was also a major expansion of eye tissue into the diencephalic region (Fig. 2a–h) in 55% (62/112) of injected embryos, with either one eye (41/62) or both (21/62) affected. The extent of this expansion varies between embryos (Fig. 2e–h), perhaps owing to mosaic distribution of the injected RNA (data not shown). Co-injection of increasing amounts of RNA encoding full-length *Sek-1*, which does not affect forebrain development when injected on its own, progressively rescued the disruption (Fig. 2k), indicating that *Sek-1* is capable of homodimerization and that the effect of truncated *Sek-1* is due to an inhibition of signal transduction.

During normal development, the presumptive eye is initially continuous with the forebrain, then separates so that it is only connected through the optic stalk¹⁶. However, in injected embryos this separation failed to occur, and the staining pattern of HNK-1 antibody suggested that there was a loss or displacement of diencephalic structures (Fig. 2e–h). Immunocytochemical analyses of the differentiated eye showed that various retinal cell types were present and correctly organized in the expanded eyes (Fig. 2i, j) and that retinal ganglion cells extended their axons (Fig. 2f), but the pigment epithelium was partially deficient (Fig. 2b, d, f–j).

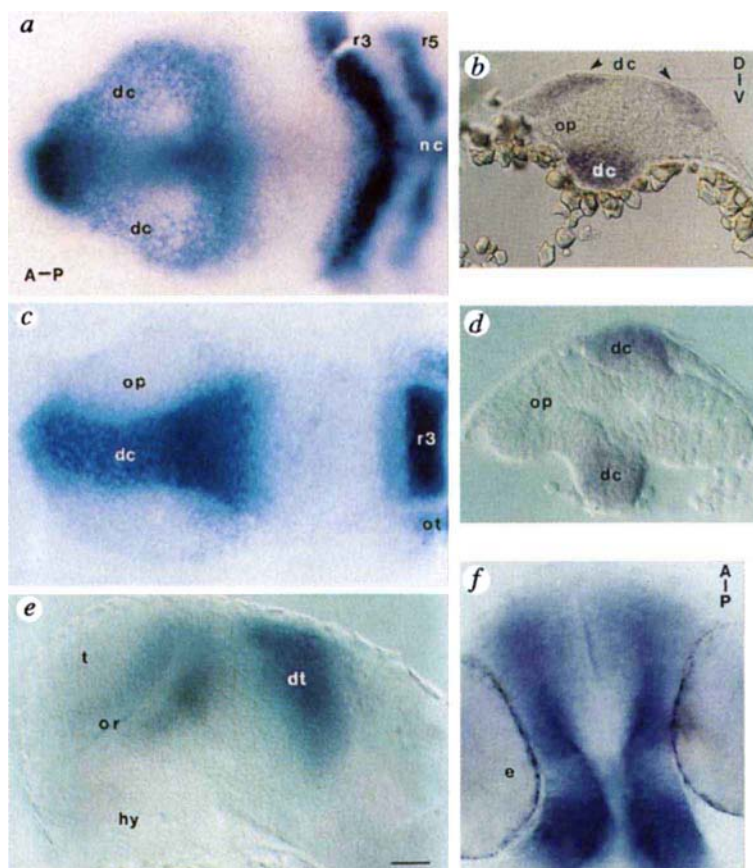


FIG. 1 Expression of *rtk1* in the developing zebrafish forebrain. *rtk1* transcripts were detected by whole-mount *in situ* hybridization¹⁵ and photographs were taken of embryos mounted under a coverslip (a, c, f, dorsal view; e, lateral view) or cross-sectioned (b, d). a, b, 10-h embryo: expression in the diencephalic primordium occurs in the ventral and dorso-lateral neuroepithelium. c, d, 13-h embryo: the dorso-lateral expression domains have converged during neural keel formation. e, f, 24-h embryo: expression has become restricted within the diencephalon to the presumptive dorsal thalamus and the region adjacent to the optic recess. *rtk1* transcripts are not detected at any stage in the optic primordium. Scale bar, 50 μ m. A–P, anterior–posterior; D–V, dorsal–ventral; dc, presumptive diencephalon; dt, dorsal thalamus; e, eye; hy, hypothalamus; nc, notochord; op, optic primordium; or, optic recess; ot, otic vesicle; r, rhombomere; t, telencephalon.

We investigated whether abnormal development of the forebrain could be detected before differentiation of the eye. Development of the vertebrate eye involves interactions between the rostral surface ectoderm, the underlying neuroectoderm and surrounding mesenchymal tissues²⁸. As *rtk1* is expressed in axial mesoderm in the head¹², it was possible that defective formation of this tissue led to disrupted forebrain patterning, but analysis with several markers, *rtk1*, *gsc* and *shh* (data not shown), argues against this possibility. In contrast, expression of *rtk1* was disrupted in the presumptive forebrain of 12-h embryos, with expressing and non-expressing cells intermingled in the prospective dorsal diencephalon, and reduced expression in the ventral diencephalon (Fig. 3a, b). At 24 h of development, the anterior *rtk1* expression domain was often absent or greatly reduced in injected embryos compared with controls (Fig. 3c, d). Similarly, expression in injected embryos of *zash-1a* (ref. 17) was decreased or absent in the developing ventral diencephalon and anterior dorsal forebrain, whereas expression in the pineal gland and ventral midbrain was not affected (Fig. 3e, f). Furthermore, *pax6*, normally expressed in the eye and dorsal diencephalon¹⁸, was ectopically expressed in some cells in the ventral diencephalon of injected 12-h embryos (Fig. 3g, h). Ectopic Pax6-positive cells persist in the presumptive

ventral diencephalon in the 18-h embryo (Fig. 3i, j), and the expansion of the retina was more evident at 24 h of development (Fig. 3k, l). Taken together, these data indicate that an alteration of regional gene expression has taken place at an early stage of development.

Cell-labelling experiments were done to clarify the fate of the disrupted diencephalic region. After injection of the lipophilic dye, DiI, into the presumptive ventral diencephalon of the 12-h or 24-h embryo, labelled cells were always incorporated into the expanded part of the retina (Fig. 4a–e). In contrast, when DiI was injected into the ventral diencephalon of control embryos, it never labelled the retina (data not shown).

The diencephalon develops from subdivisions that, based upon gene expression patterns, are specified at early neurula stages^{1,19}. However, regionalization of the neural ectoderm is gradual²⁰ and at early stages these subdivisions are not committed to a specific identity, as shown, for example, in the chick by grafts of mesencephalic tissue that induce a mesencephalic phenotype in the adjacent diencephalon^{21–24}. This plasticity is consistent with the fate map of 10-h zebrafish neuroectoderm in which the retina and diencephalon progenitors overlap in the border region²⁵ at the stage that *rtk1* is expressed in the presumptive diencephalon.

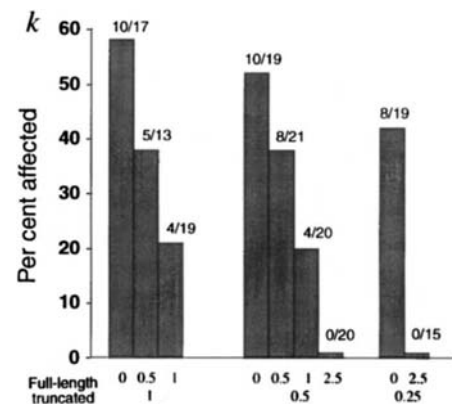
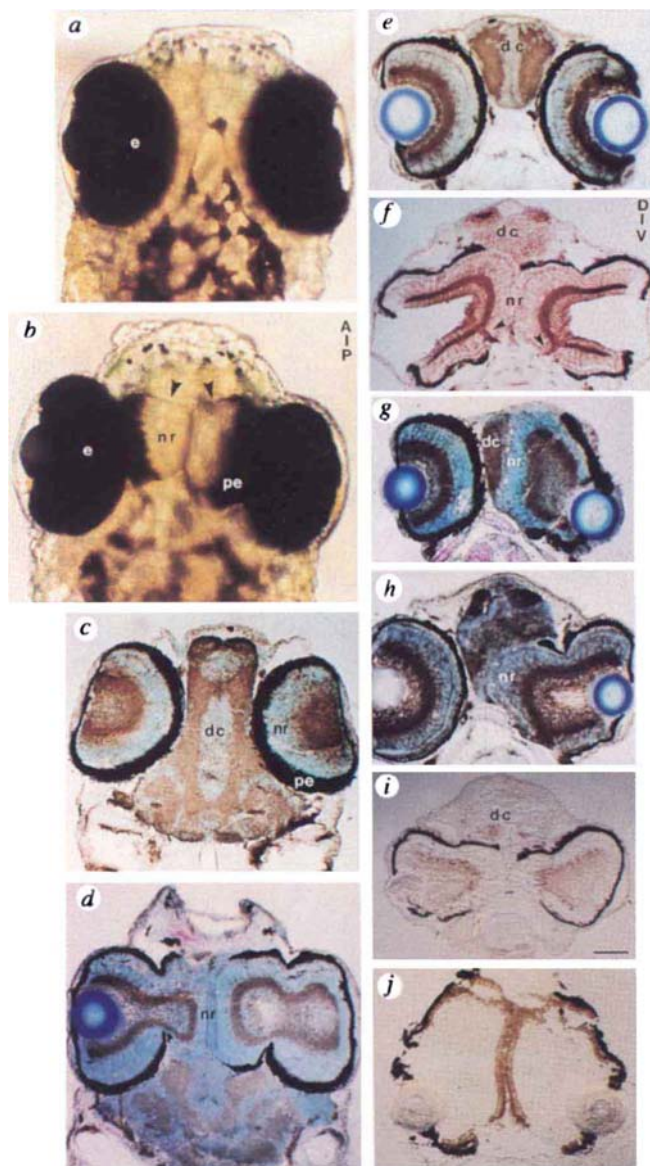


FIG. 2 Effect of truncated Sek-1 on forebrain development in the zebrafish embryo. Zebrafish embryos were injected with RNA encoding truncated Sek-1 and analysed at 5 days of development. a, b, Dorsal view of whole-mount preparation. The eye has expanded towards the midline in injected (b) compared with control (a) embryos. c–h, Immunocytochemistry with HNK-1 antibody, detecting axon tracts and neuronal cell bodies, was carried out on frontal (c, d) or transverse (e–h) sections through the diencephalon. Varied expansions of the retina occurred in injected (d, f–h) compared with control (c, e) embryos, with the expanded retina replacing the ventral diencephalon (f), half of the entire diencephalon (g), or encroaching across the midline to the other side of the ventral diencephalon (h). The pigment epithelial layer was often absent next to the expanded part of the retina (d, f–h), and unlike in control embryos, the retina has failed to separate from the diencephalon (g, h). i, j, Retinal differentiation in injected embryos was examined by immunocytochemistry of transverse sections with Fret-43 antibody (i) or anti-Pax6 antibody (j). These stain cone photoreceptors and amacrine cells, respectively, which had a normal organization even in the absence of pigment epithelial cells. k, Rescue of disruption to eye development by full-length Sek-1. RNA encoding truncated Sek-1 was co-injected with various amounts of RNA encoding full-length Sek-1. The embryos were then analysed at 5 days of development by scoring morphological defects in the eye; each bar represents the percentage of embryos affected, with the number of affected/total indicated above. The amounts of RNA are normalized relative to the maximum amount (0.3 ng) of truncated receptor RNA injected. Large arrowheads indicate the abnormal eye, small arrowheads indicate optic nerve fibres, and asterisks indicate the border between the ectopic retina and the diencephalon. Scale bar, 50 μ m. dc, Diencephalon; nr, neural retina; pe, pigment epithelium. Other labels as for Fig. 1.

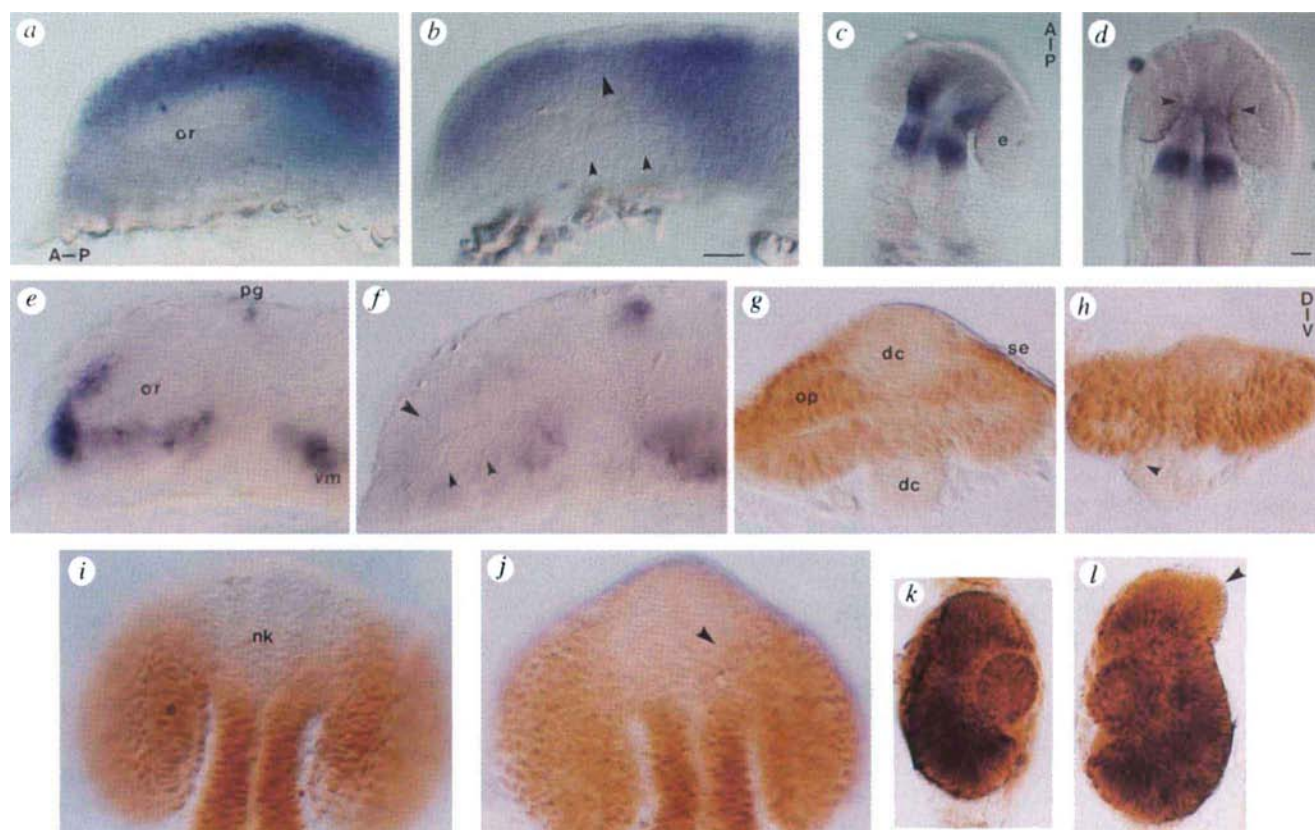
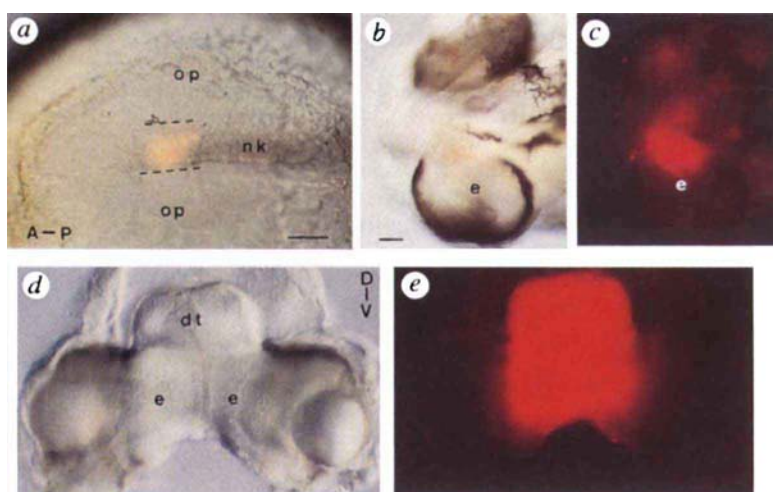


FIG. 3 Effect of truncated Sek-1 on gene expression in the developing rostral CNS. Embryos injected with RNA encoding truncated Sek-1 were fixed at various stages and analysed by whole-mount *in situ* hybridization or immunohistochemistry. *a–d*, Probed with *rtk1* and shown in lateral (*a, b*) and dorsal view (*c, d*). At 12 h of development, in control embryos (*a*) domains of *rtk1* expression occur in the diencephalon, at a higher level in the dorsal and lower level in the ventral region. In the injected embryo shown in *b*, the ventral domain of expression was reduced and there are non-expressing cells present in the dorsal region. At 24 h, the anterior expression domain is greatly reduced in the injected embryo (*d*) compared with the control (*c*). *e, f*, Lateral view of 12-h embryos probed with *zash-1a*. This gene is expressed in discrete groups of cells in the presumptive diencephalon, ventral mesencephalon and pineal gland (*e*). The level of

zash-1a expression is greatly decreased in the forebrain region but is maintained in the presumptive pineal gland and ventral midbrain (*f*). *g–l*, Control (*g, i, k*) or injected (*h, j, l*) embryos stained with Pax6 antibody. *g, h*, Transverse sections of a 12-h embryo. In control embryos (*g*), Pax6-expressing cells are not observed in the ventral diencephalon, whereas expressing cells are observed in this region of the injected embryo (*h*); several ectopically expressing cells are present, but only one is seen in this focal plane. *i, j*, Dorsal view of 18-h embryos. Note a population of ectopic Pax6-expressing cells adjacent to the retina (*j*). *k, l*, Dissected eye from a 24-h embryo. The eye is enlarged in the injected embryo (*l*). Arrowhead indicates the affected area. Scale bar, 50 μ m. nk, Neural keel; pg, pineal gland; se, surface ectoderm; vm, ventral mesencephalon. Other labels as for previous figures.

FIG. 4 Dil labelling of embryos expressing truncated Sek-1. Embryos injected with RNA encoding truncated Sek-1 or control embryos were labelled by injection of Dil into the presumptive ventral diencephalon. *a–c*, Iontophoretic injection of Dil into the presumptive ventral diencephalon (part of the neural keel) of 12-h embryo, visualized immediately after injection (*a*, double exposure photograph), or at 3 days (*b*, bright field; *c*, epifluorescence photograph). These injections routinely labelled 5–10 cells in the neural keel, and there was no labelling of the presumptive eye at 12 h. Dotted lines indicate the edge of the neural keel. However, in 3-day embryos, the labelled cells have contributed to the expanded part of the eye. *d, e*, Transverse sections of 3 day embryo after pressure injection of Dil into the presumptive ventral diencephalon of a 24 h embryo, viewed under bright field (*d*) and epifluorescent illumination using a rhodamine filter (*e*). Note that Dil is only present in the expanded part of the eye while the original part of the eye is devoid of label. Scale bars, 50 μ m. Labels as for previous figures.

METHODS. Iontophoretic injection of Dil to label small groups of cells was carried out on 12 h embryos after dechorionation and orientating in 1.2% agarose in 10% Hank's. Microelectrodes with a 1–1.5- μ m tip were filled at the tip with 3 mg ml⁻¹ Dil in dimethylformamide, back-filled with 1M LiCl, and injection carried out by applying 9 V for 5 s. In



24-h embryos, larger groups of cells were labelled by pressure injection of 3 mg/ml Dil in mineral oil into the ventral diencephalon, either into control embryos or the constricted area of affected embryos.

Three models could explain the expansion of eye territory after inhibition of *rtk1* function. First, the ectopic eye could derive entirely from presumptive retinal cells that have migrated into presumptive diencephalon because of an abnormal intermingling between territories. Second, it could derive entirely from presumptive diencephalic cells that have switched to a retinal identity. Third, it might derive from both sources, because intermingling between presumptive retina and diencephalon has caused the latter to switch to a retinal identity as a result of homogenetic induction. Our data argue against the first model, because it does not account for the lack of expression of molecular markers of the ventral diencephalon. The second and third models both involve a switch of presumptive diencephalon to a retinal identity, but differ in whether *rtk1* activation is directly required in diencephalic cells to maintain their identity, or whether the switch is secondary to disruption of a restriction of cell movement. The altered gene-expression patterns at 12 h of development argue that such effects on cell identity and/or increased intermingling have already occurred by this stage. As the effects of truncated Sek-1 on the hindbrain also indicate potential roles in regulating cell identity or in restricting cell movement¹⁰, it is tempting to suggest a similar function in these different regions of the developing brain. A critical question is the identity and expression pattern of the ligand for *rtk1* in the diencephalon. The ELF-1 ligand that binds Sek-1 (ref. 26) is not expressed in the diencephalon or eye, so another ligand(s) must interact with *rtk1* in this region. It is intriguing that in the retinotectal projection⁸ and the hindbrain²⁷, Eph-related RTKs and their ligands are expressed in complementary domains, consistent with roles in the repulsion of axons⁹ or cells¹⁰. It will therefore be interesting to see whether a similar situation occurs in the zebrafish diencephalon, and to analyse the

function of other members of the Eph-related RTK family that could cooperate in the patterning of this region. □

Received 27 December 1995; accepted 3 April 1996.

1. Rubenstein, J. L. R. & Puelles, L. in *Current Topics in Developmental Biology* Vol. 29 (ed. Pedersen, R. A.) 1–63 (Academic, London, 1994).
2. Macdonald, R. et al. *Neuron* **13**, 1039–1053 (1994).
3. van der Geer, P., Hunter, T. & Lindberg, R. A. *Rev. Cell Biol.* **10**, 251–337 (1994).
4. Lai, C. & Lemke, G. *Neuron* **6**, 691–704 (1991).
5. Pasquale, E. B. et al. *J. Neurosci.* **12**, 3956–3967 (1992).
6. Nieto, M. A. et al. *Development* **116**, 1137–1150 (1992).
7. Pandey, A., Lindberg, R. A. & Dixit, V. M. *Curr. Biol.* **5**, 986–989 (1995).
8. Cheng, H.-J. et al. *Cell* **82**, 371–381 (1995).
9. Drescher, U. et al. *Cell* **82**, 359–370 (1995).
10. Xu, Q. et al. *Development* **121**, 4005–4016 (1995).
11. Irving, C. et al. *Dev. Biol.* **173**, 26–38 (1996).
12. Xu, Q. et al. *Development* **120**, 287–299 (1994).
13. Hemmati-Brivanlou, A. & Melton, D. A. *Nature* **359**, 609–614 (1992).
14. Graff, J. M. et al. *Cell* **79**, 169–179 (1994).
15. Amaya, E., Musci, T. J. & Kirschner, M. W. *Cell* **66**, 256–270 (1991).
16. Schmitt, E. A. & Dowling, J. E. *J. comp. Neurol.* **344**, 532–542 (1994).
17. Allende, M. L. & Weinberg, E. S. *Dev. Biol.* **166**, 509–530 (1994).
18. Krauss, S. et al. *EMBO J.* **10**, 3609–3619 (1991).
19. Figdor, M. C. & Stem, C. D. *Nature* **363**, 630–634 (1993).
20. Papalopulu, N. *Perspect. dev. Neurobiol.* **3**, 39–52 (1995).
21. Chung, S. H. & Cooke, J. *Nature* **258**, 126–132 (1975).
22. Nakamura, H. et al. *Cell Diff. Dev.* **19**, 187–193 (1986).
23. Alvarado-Mallart, R., Martinez, S. & Lance-Jones, C. *Dev. Biol.* **139**, 75–88 (1990).
24. Martinez, S., Wassef, M. & Alvarado-Mallart, R. M. *Neuron* **6**, 971–981 (1991).
25. Woo, K. & Fraser, S. E. *Development* **121**, 2595–2609 (1995).
26. Cheng, H.-J. & Flanagan, J. G. *Cell* **79**, 157–168 (1994).
27. Bergemann, A. D. et al. *Molec. cell. Biol.* **15**, 4921–4929 (1995).
28. Macdonald, R. et al. *Development* **121**, 3267–3277 (1995).

ACKNOWLEDGEMENTS. This work was supported by the MRC. We are grateful to our colleagues for many useful discussions, S. Wilson for his critical reading of the manuscript, J. Clark for help with iontophoretic labelling with Dil, E. Weinberg for *zash-1a* probe, and B. Gasking for care of the fish. N.H. is a BBSRC Senior Fellow.

CORRESPONDENCE and requests for materials should be addressed to N.H. (e-mail: udb1309@bay.cc.kcl.ac.uk).

Multiple spike-initiation zones in single neurons revealed by voltage-sensitive dyes

Dejan Zečević

Department of Cellular and Molecular Physiology,
Yale University School of Medicine, 333 Cedar Street, New Haven,
Connecticut 06520, USA

THE primary function of the nerve cell is to process electrical signals. Over the past fifteen years¹ there has been renewed interest in the detailed spatial analysis of signalling in individual neurons owing to experimental evidence that the regional electrical properties of neurons are complex. Thus the behaviour of many nerve cells cannot be understood on the basis of micro-electrode measurements from the soma. Regional electrical properties of neurons have been studied using sharp microelectrode¹ and patch-electrode^{2,3} recordings from neuronal processes, high-resolution multisite optical recordings of Ca²⁺ concentration changes^{4,5}, and by using models to predict the distribution of membrane potential in the entire neuronal arborization⁶. Additional direct evidence about electrical signalling in neuronal processes *in situ* can now be obtained by recording membrane potential changes using voltage-sensitive dyes^{7,21}. Here I demonstrate the existence of multiple action potential trigger zones in separate regions of the neuronal arborization of an identified molluscan neuron.

Experiments were carried out on metacerebral giant cells (MGC)⁸ within the cerebral ganglia of the land snail *Helix aspersa*. Fig. 1A shows a cell selectively stained by microinjection with a positively charged, membrane-impermeant, voltage-sensitive, fluorescent, amino-naphthyl styryl dye known as JPW1114

(ref. 7). There are two symmetrical cells with the same branching pattern in every preparation.

It was not possible to detect any pharmacological effects or photodynamic damage caused by the dye. The absence of large pharmacological effects is evident from the fact that electrically recorded action potentials were essentially unchanged after staining and an incubation period of more than 12 hours (Fig. 1B, *a, b*). Photodynamic damage was likely to be either absent or insignificant as the time course of the electrically recorded action potential from the cell body and optically recorded signals from the soma and neuronal processes did not change over several averaging trials. This result demonstrates that both soma and neuronal processes are apparently insensitive to photodynamic damage under these experimental conditions. However, it is not possible, at present, to be absolutely certain that staining or photodynamic damage does not influence regional properties of the neuron in some subtle way, because in the absence of voltage-sensitive dye, no independent measurements with adequate spatial resolution are available. All that can be said is that no clear evidence for such effects was found.

Near-threshold electrical stimulation of the soma evoked action potentials that were recorded by a microelectrode with a delay (Fig. 1B, *a, b*), indicating that the site of origin of the spike might be remote from the recording location. The position of the spike trigger zone cannot, however, be determined from a single microelectrode recording.

Multisite optical recording of membrane potential changes revealed that several independent action potentials are initiated at separate locations in the axonal branching structure in response to direct soma stimulation. Figure 2 illustrates the result of optical recording from axonal branches Br1 and Br2 during the time the neuron was stimulated by passing current through a microelectrode in the soma. Two separate action potentials were recorded: one dominating the signal at location 1 and the other dominating the signal at location 3. The membrane potential between locations 1 and 3 was influenced, to different degrees, by both

FIG. 1 A, Giant metacerebral neuron from the left cerebral ganglion following 12 h incubation at 15 °C after injection with the fluorescent voltage-sensitive dye JPW1114, whose structure is shown. The cell body and main processes are close to the surface of the ganglion and clearly visible in the unfixed preparation. Axonal branches are marked Br1–5. Br1 projects to the cerebral commissure. Br2 divides within a ganglion into Br3, which projects to the external lip nerve, and Br4, which projects to the cerebral–buccal connective. Br5 projects to the internal lip nerve. Excitation: 540 ± 30 nm; dichroic mirror: 570 nm, barrier filter: 610 nm. The dissection procedure, pressure injection of the dye and staining protocol were done as described⁷. B, Standard microelectrode recordings of action potentials initiated in MGCs by passing current (indicated by the bar) through the recording electrode in the soma (a, b) or by stimulating the peripheral nerve (Br3) that contains the axon of the cell (c); a and b are recordings from the same cell before and after dye injection and 12 h incubation. No pharmacological effects were detected. c, Recording of a spike evoked antidromically, by stimulating Br3 through a suction electrode attached to a cut end of a nerve, in a cell injected with the dye and incubated at 18 °C for 22 h. Action potential propagation along the process was not affected by staining.

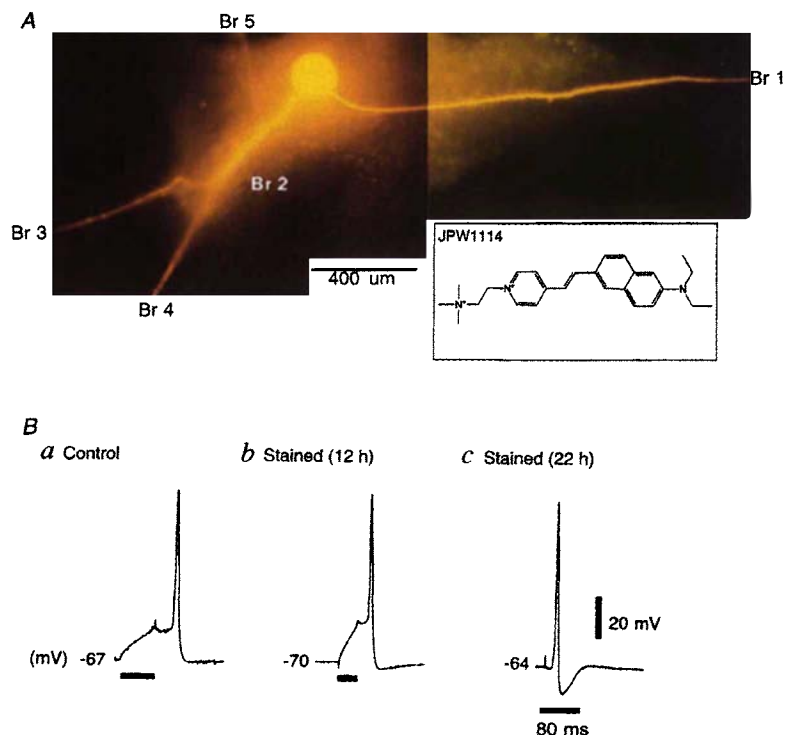
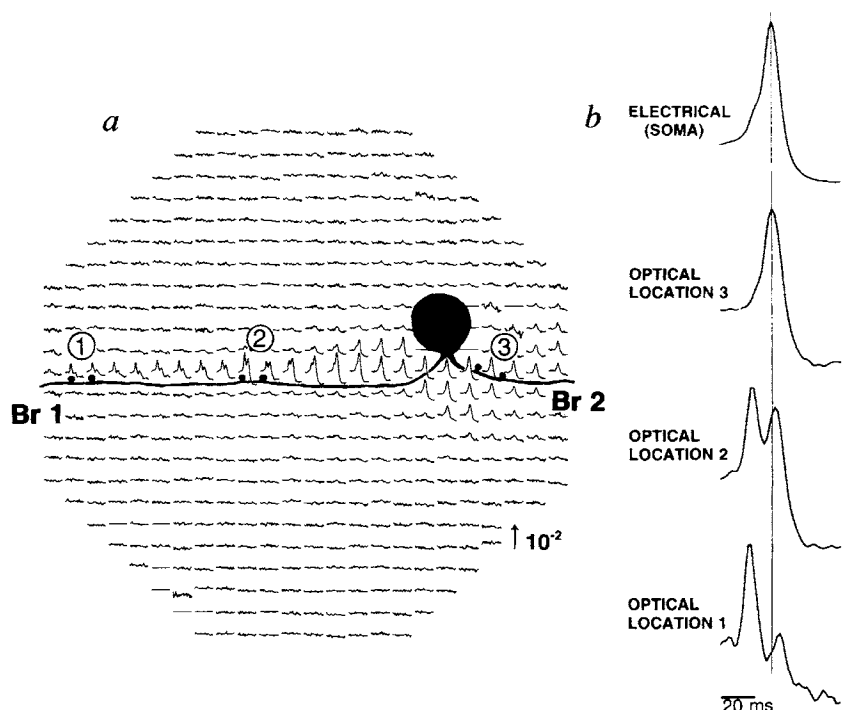


FIG. 2 Optical recordings of action potential signals from elements of a photodiode array along the enlarged and inverted image of a metacerebral cell from the right cerebral ganglion. Each diode received light from a 50×50 - μ m area in the object plane. Each trace represents the output of one diode for 70 ms when the giant metacerebral cell was stimulated to produce an action potential by passing current through a microelectrode in the soma. The recordings are divided by the resting light intensity. Arrow indicates the direction and stated value of the relative fluorescence change (dF/F). The cell was depolarized from the resting membrane potential of -68 mV. In all of the experiments designed to determine the location of the primary spike trigger zone, near-threshold stimulus was used. Twenty four trials were averaged. Optical signals corresponding to evoked action potentials were clearly recorded from both soma (the largest signals from the soma are eliminated from the figure for clarity) and processes. The cell body and axonal branch Br2 were slightly out of focus; consequently, signals from the soma and Br2 are spread over several detectors. The interaction of two spikes initiated by the same stimulus at two separate trigger zones in Br1 and Br2 is shown in a. Note that the electrical recording from the soma (b), does not contain information about spatial aspects of signal interactions in the axon.



METHODS. The fluorescent image of the cell was projected onto the 464-element photodiode array (Centronix) by an objective (Wild, $10\times$ 0.4NA). A 150-W xenon arc lamp (model 1600XT/A, Opti-Quip) was used as a light source. The best signals were obtained by using an excitation interference filter at 520 ± 45 nm, a 570-nm dichroic mirror and a 610-nm barrier filter (Schott RG610). For many voltage-sensitive dyes, absorption and fluorescence changes are both fast and linear with membrane potential changes in the squid giant axon^{24–26}. This was confirmed on a *Helix* metacerebral interneuron by comparing electrical signals with optically recorded action potentials⁷. Optical signals associated with action potentials, expressed as fractional changes in fluorescent light intensity (dF/F), were about 1% in recordings from neuronal processes. The signal size dropped abruptly in the most distal axonal regions (Br1 in a and Br3 in Fig. 3a) because the concentration of dye declines with the distance from the soma, and so the resting light for the detectors that receive light from

distal processes becomes dominated by autofluorescence of the preparation. Slow changes in light intensity due to bleaching of the dye (decaying baseline preceding spikes in recordings near the soma in Fig. 3a) were about 100 times slower than the rising phase of the action potential and had no effect on timing information on which my conclusions were based. This was confirmed by correcting for the bleaching effect in the exemplar traces by subtracting an appropriate exponential function (not shown). The output of each diode was converted to voltage, filtered, digitized (adjustable frame rate from 0.6 ms per full frame) and stored in the computer that was also used to analyse and display data⁷.

impulses; the existence of the two action potentials is obvious at location 2. At least two separate spike trigger zones must exist to account for the recorded collision of impulses. A comparison of the optical signals and the simultaneously recorded electrical signal from the soma (first trace in Fig. 2b) shows clearly that the information about interaction of signals in specific regions in the processes cannot be recovered from the microelectrode recording.

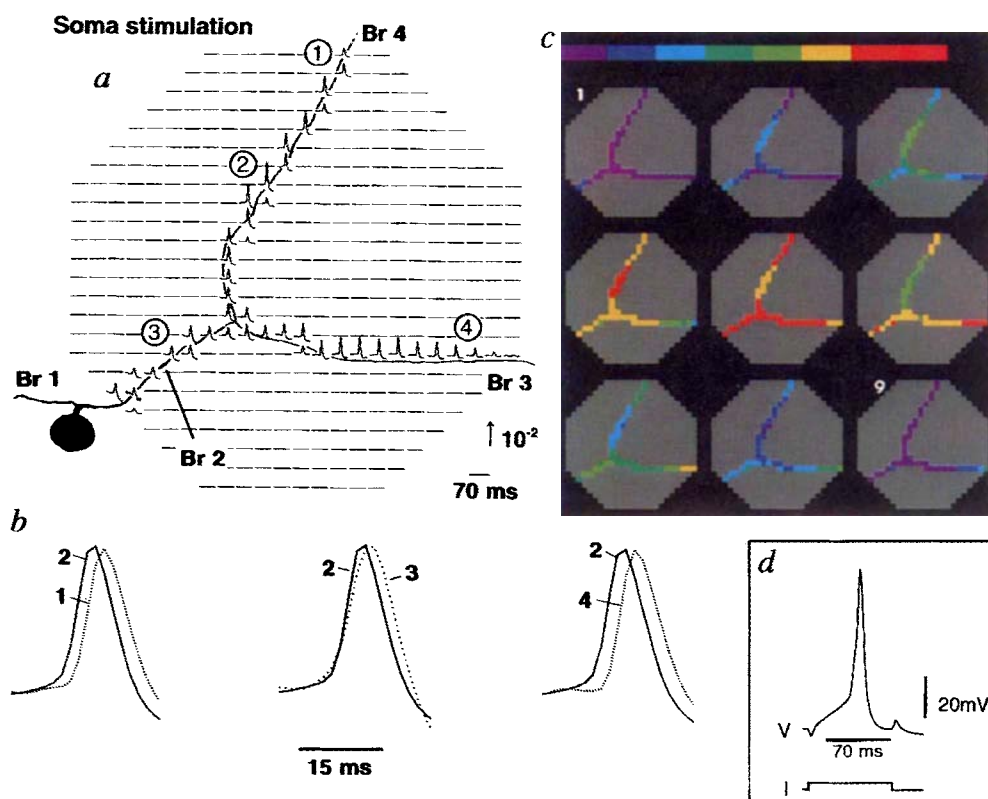
In eleven additional preparations, the direction of action potential propagation in axonal branches Br1, Br2 and Br5 was determined by comparing optical recordings from different locations. It was invariably found that the first spike in Br1 was initiated in a distal axonal segment, greater than 800 μm from the soma (as in Fig. 2) and propagated antidromically towards the cell body. Separate action potentials were also initiated by the same stimulus in branches Br2 and Br5 (not shown) and also propagated antidromically towards the soma from the distal site of origin.

These results show that there must be at least three separate spike trigger zones in the branching structure of the metacerebral neuron, all of which are far from the soma. Although the experiment shown in Fig. 2 does not indicate the actual site of either trigger zone, the data indicate clearly that they must be positioned at or distal to locations 1 and 3. To understand the functional organization of this neuron, experiments were carried out to determine the position of the trigger zones in different processes. There were a number of difficulties. First, axonal branches become thinner with distance from the soma. Second, the concentration of the intracellular dye declines with distance from the site of injection (soma). Because the signal-to-noise ratio in optical recording is proportional to the light intensity and hence, in fluorescence measurements, to both membrane area

imaged and dye concentration, it was not possible to follow signals, with present sensitivity, sufficiently far to determine the position of distal trigger zones in branches Br1 and Br5.

It was possible, however, to investigate spike initiation in the branching structure of Br2. It was consistently found, in seven different preparations, that the first spike originated from a distal part of Br4. In repeated measurements from two of these preparations, the signal-to-noise ratio was good enough to locate the spike trigger zone. Figure 3a represents multisite optical recording of action potential signals, evoked by a transmembrane current pulse (Fig. 3d), from axonal branches Br2, Br3 and Br4. In Fig. 3b, recordings from different locations, scaled to the same height, are compared to determine the site of origin of the action potential. It appears that the earliest action potential, in response to soma stimulation, was generated near location 2, in the axonal branch Br4, situated in the cerebral–buccal connective outside the ganglion. The spike propagated orthodromically from the site of initiation toward the periphery in branch Br4, and antidromically toward the soma and into the branch Br3 in the external lip nerve. The antidromic spike caused the soma to discharge. The spike then propagated into the main axonal branch Br1, where collision with another action potential, evoked separately in a distal region of Br1 by the same depolarizing stimulus, was detected (not shown). Although the difference in the timing of the spikes at locations 2 and 3 (Fig. 3b) is small, the direction of propagation is clear from the colour-coded representation of the data shown in Fig. 3c, in which the potential changes in the branching structure are indicated at nine different times; red corresponds to the peak of the action potential. The panels show the position of the action potential trigger zone at location 2 and ortho- and antidromic spread of the nerve impulse from the site of initiation. The earliest spike was evoked ~ 1 mm away from the soma. The spike initiation

FIG. 3 a, Optical recordings of action potential signals from elements of the photodiode array positioned over the image of axonal arborization of a metacerebral cell from the left cerebral ganglion. Spikes were evoked by transmembrane current pulses, as shown in d, delivered through the recording microelectrode in the soma. The cell was depolarized from the resting membrane potential of -64 mV. Each optical trace in a represents 70 ms of recording centred around the peak of the spike as indicated by the bar in d. Recordings are divided by the resting light intensity. The arrow indicates the direction and the stated value of the relative fluorescence change (dF/F). Each diode received light from $50 \times 50 \mu\text{m}$ area in the object plane. Ninety trials were averaged to improve the signal-to-noise ratio from distal processes. Optical signals were found in the regions of the array that correspond closely to the geometry of the cell. Data from the detectors that do not receive light from the cell were deleted to improve clarity and allow colour-coded display shown in c, based on a relative scale applied separately to each trace. b, Recordings from 4 different locations indicated in a, scaled to the same height, are compared to determine the site of the origin of the action potential and the direction of propagation. c, Colour-coded representation of the data shown in a, indicating the size and location of the primary spike trigger zone and the pattern of spike propagation. Consecutive frames represent data points that are 1.6 ms apart. Colour scale is in relative units,



with the peak of the action potential shown in red. d, Microelectrode recording of the action potential from the cell body (V) evoked by transmembrane current pulse (I). Note that different timescales were used for the optical (b) and electrical traces (d).

segment in the axon is roughly 300 μm in length. Under normal conditions, slow depolarizing voltage pulses applied to the soma are electronically spread into the processes with little attenuation, owing to the very long axonal space constant for slow voltage changes^{3,9}. These depolarizing pulses initiate action potentials at remote sites in the processes that are characterized by a higher excitability than that of neighbouring segments. The existence of multiple and remote trigger zones has been previously inferred from the correlation of extra- and intracellular electrical recordings in several neuronal types^{10–13}.

The results shown in Fig. 3 provide information about the spike trigger zone in a neuron with complex geometry. This kind of information could constrain the choice of channel densities and geometrical factors that are used in biophysical models of the trigger zone in neurons¹⁴.

In conclusion, optical recording revealed the existence of multiple spike trigger zones in separate axonal branches of a neuron. The fact that membrane excitability appears to be different in different regions of the same neuron is well established^{1,3,9,15–17}. Also, the existence of multiple spike initiation sites in *Aplysia* neurons led to the important postulate¹⁰ that neurons might be functionally subdivided. A consequence of this would be that spike initiation and the pattern and nature of signal processing might be complex and impossible to predict in the absence of detailed measurements. Optical recording^{18–20} from individual neurons^{7,15,21} allows this kind of measurement in favourable preparations. Relatively modest improvements in signal-to-noise ratio and in dye characteristics, such as the rate of dye spread within a neuron²², would make similar studies possible in vertebrate neurons, including direct mapping of synaptic inputs onto the neuronal arborization. In initial experiments on rat cerebellar slices,

optical signals corresponding to membrane potential transients were recorded from Purkinje neurons using the same dye²³. $\square$

Received 8 September 1995; accepted 20 March 1996.

1. Linas, R. & Sugimori, M. *J. Physiol.* **305**, 197–213 (1980).
2. Stuart, G. J. & Sakmann, B. *Nature* **367**, 69–72 (1994).
3. Stuart, G. J. & Häusser, M. *Neuron* **13**, 703–712 (1994).
4. Ross, W. N. et al. *Jap. J. Physiol.* **43**, S83–S89 (1993).
5. Eilers, J., Augustine, G. J. & Konnerth, A. *Nature* **373**, 155–158 (1994).
6. Traub, R. D., Jefferys, J. G. R., Miles, R., Whittington, M. A. & Tóth, K. J. *Physiol.* **481**, 79–95 (1994).
7. Antić, S. & Zečević, D. *J. Neurosci.* **15**, 1392–1405 (1995).
8. Kandel, E. R. & Tauc, L. *J. Physiol.* **183**, 269–286 (1965).
9. Tauc, L. *J. gen. Physiol.* **45**, 1077–1097 (1962).
10. Tauc, L. & Hughes, G. H. *J. gen. Physiol.* **46**, 533–549 (1963).
11. Korn, H. & Bennett, M. V. L. *Brain Res.* **27**, 169–175 (1971).
12. Heitler, W. & Goodman, C. S. *J. exp. Biol.* **76**, 63–84 (1978).
13. Combes, D., Simmers, J., Nonnotte, L. & Moulins, M. *J. Physiol.* **460**, 581–602 (1993).
14. Huguenard, J. R., Joerges, J., Mainen, Z. F. & Sejnowski, T. J. *Soc. Neurosci. abstr.* **20**, 1527 (1994).
15. Ross, W. N. & Krauthamer, V. J. *Neurosci.* **4**, 659–672 (1984).
16. Midtgard, J., Lasser-Ross, N. & Ross, W. N. *J. Neurophysiol.* **70**, 2455–2469 (1993).
17. Regehr, W. G., Kehoe, J. S., Ascher, P. & Armstrong, C. *Neuron* **11**, 145–151 (1993).
18. Cohen, L. B. & Salzberg, B. *Rev. Physiol. biochem. Pharmac.* **83**, 35–88 (1978).
19. Wu, J. Y. & Cohen, L. B. in *Fluorescent and Luminescent Probes for Biological Activity* (ed. Mason, W. T.) 389–404 (Academic, New York, 1993).
20. Zečević, D. et al. *J. Neurosci.* **9**, 3681–3689 (1989).
21. Grinvald, A., Salzberg, B. M., Lev-Ram, V. & Hildesheim, R. *Biophys. J.* **51**, 643–651 (1987).
22. Grinvald, A., Lieke, E. E., Frostig, R. D. & Hildesheim, R. *J. Neurosci.* **14**, 2545–2568 (1994).
23. Kogan, A., Ross, W., Zečević, D. & Lasser-Ross, N. *Brain Res.* **700**, 235–239 (1995).
24. Gupta, R. K. et al. *J. Membr. Biol.* **58**, 23–137 (1981).
25. Cohen, L. B. & Leshner, S. *Soc. gen. Physiol. ser.* **40**, 71–99 (1986).
26. Salzberg, B. M., Obaid, A. L. & Bezanilla, F. *Jap. J. Physiol.* **43**, S37–S41 (1993).

ACKNOWLEDGEMENTS. I thank L. Loew and J. Wuskul for dyes; L. Cohen, K. Chandler, W. Ross and B. Salzberg for comments on the manuscript; and A. Cohen for data acquisition and analysis software; and J. Mijović and S. Pejanović for discussion at Smederevo. This work was supported by a grant from the NIH.

CORRESPONDENCE and requests for materials should be addressed to the author (e-mail: Zec@biomed.med.yale.edu).

Regulation of B-lymphocyte negative and positive selection by tyrosine phosphatase CD45

Jason G. Cyster*, James I. Healy*, Kenji Kishihara†, Tak W. Mak‡, Matthew L. Thomas§ & Christopher C. Goodnow*

* Department of Microbiology and Immunology, Program in Immunology and Howard Hughes Medical Institute, Beckman Center, Stanford University School of Medicine, Stanford, California 94305, USA

† Department of Immunology, Medical Research Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812, Japan

‡ The Amgen Institute and Ontario Cancer Institute, Department of Medical Biophysics and Immunology, University of Toronto, Toronto, Ontario, Canada M4X 1K9

§ Howard Hughes Medical Institute, Department of Pathology, Washington University School of Medicine, St Louis, Missouri 63110, USA

ELIMINATION of self-reactive B cells must be balanced against the need for B-cell diversity for antibody responses to pathogens. To analyse factors that determine the extent of B-cell negative selection, we crossed CD45-deficient mice¹ with mice carrying immunoglobulin transgenes specific for hen egg lysozyme (HEL). CD45 positively regulates antigen-receptor signalling^{2–9} and CD45-deficient HEL-specific B cells gave diminished signalling in response to HEL. Significantly, few mature CD45^{−/−} B cells accumulated, despite normal immature B-cell production. Circulating HEL autoantigen mediates negative selection of mature CD45^{−/−} HEL-binding B cells¹⁰ but, in striking contrast, the

autoantigen positively selected CD45^{−/−} HEL-binding B cells, promoting their accumulation as long-lived IgD^{hi} cells. These findings are consistent with a signal-threshold model for B-cell selection and demonstrate that changes in antigen receptor signalling can cause high-affinity self-reactive B cells to be actively retained instead of eliminated, thus revealing a potential mechanism for inherited susceptibility to autoimmune disease.

Mice with a targeted disruption of CD45 (ref. 1) were bred with mice expressing immunoglobulin heavy- and light-chain transgenes specific for HEL, in which most B cells bind HEL with uniformly high affinity¹¹. To test the ability of CD45^{−/−} B-cells to respond to antigen, splenic B cells were stimulated *in vitro* with soluble HEL. Activation of the ERK/RSK/EGFR1 signalling pathway¹² was diminished in CD45-deficient B cells (Fig. 1a,b). Similarly, the rise in intracellular calcium was reduced compared to CD45^{+/+} or CD45^{+/-} controls (Fig. 1c,d). In response to more potent receptor crosslinking by antibodies against IgM, increased ERK activation was detected (Fig. 1a) and the size of the calcium response by CD45^{−/−} B cells was similar to controls (Fig. 1c), although the maximum response was delayed (Fig. 1c, d). The early-activation marker CD86 (B7.2/B70) was induced in CD45^{−/−} cells, with a marked shift in the antigen dose-response curve and reduced maximum expression (Fig. 1e). CD54 (ICAM-1), a surface antigen expressed later after activation, was not upregulated in CD45-deficient cells (Fig. 1e). Similarly, the mutant cells failed to enlarge even at very high concentrations of soluble antigen, whereas CD45^{+/+} and CD45^{+/-} cells showed the enlargement characteristic of entry into G1 phase of the cell cycle (Fig. 1e). Direct stimulation of downstream signalling with phorbol ester and ionomycin induced normal levels of CD86 and CD54, and cell enlargement (Fig. 1e). Collectively, these findings indicate that expression of CD45 is required for efficient signalling in response to antigen.

To investigate the requirement of CD45 in B-cell selection, chimaeric mice were generated by using CD45-deficient or wild-

| Present address: Department of Microbiology and Immunology, University of California, San Francisco, California 94143-0414, USA.

TABLE 1 Maturation of B lymphocytes lacking CD45 is promoted by binding soluble HEL autoantigen

Bone marrow donors*	HEL-binding B cells gated†	Recipient:	Bone marrow ($\times 10^{-5}$)‡		Spleen ($\times 10^{-6}$)	
			Non	HEL	Non	HEL
CD45 ^{-/-} (n = 7)	All IgD ^{hi} subset		16.4 (6.0)	12.7 (3.0)	15.8 (7.5) 6.4 (4.3)	32.5 (11.1) 26.3 (7.9)
$P < 0.001$						
CD45 ^{+/+} (n = 4)	All IgD ^{hi} subset		12.0 (5.8)	14.7 (4.7)	42.4 (14.5) 24.2 (6.7)	68.4 (16.9) 45.6 (13.1)
CD45 ^{+/+} (n = 5)	All IgD ^{hi} subset		10.4 (5.6)	8.3 (1.4)	24.8 (11.2) 15.3 (2.2)	25.2 (9.4) 22.8 (7.3)

* Haemopoietic radiation chimaeras were generated as before¹⁵ and were reconstituted for 5–10 weeks. No differences were observed after different periods of reconstitution.

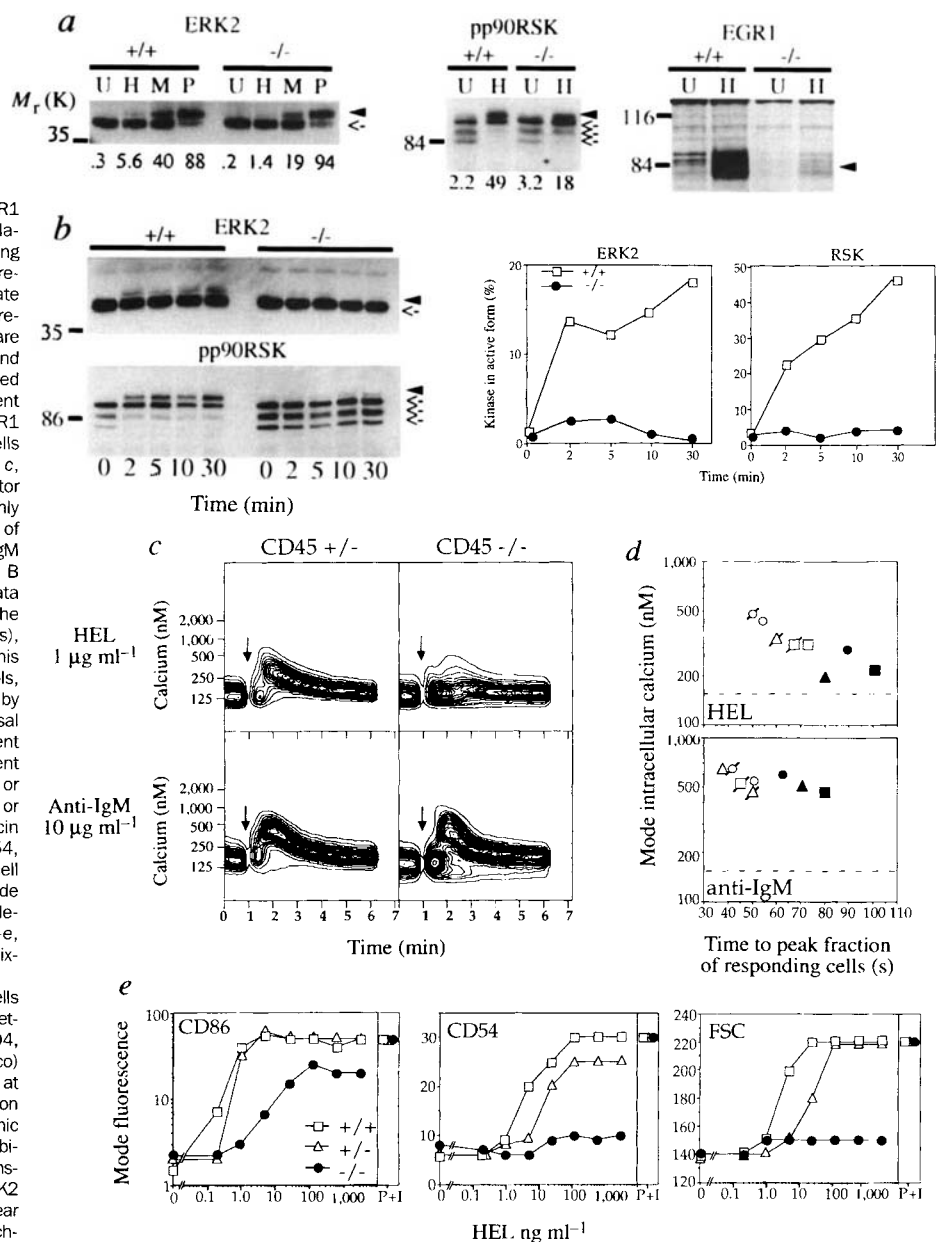
† B cells were stained and gated as for Fig. 2.

‡ Number of immature IgM⁻ IgD⁻ cells from one tibia and fibula. Data shown are arithmetic means with standard deviations in parentheses, and are from the indicated number (n) of mice per group. Statistical comparisons are by Student's *t*-test.

FIG. 1 Depressed antigen-receptor signalling in CD45-deficient B cells. **a**, Western blot analysis of phosphorylation of ERK2 (pp42, mitogen-activated protein kinase) and its substrate pp90RSK, and induction of the ERK-responsive nucleoprotein EGR1 in purified CD45^{+/+} and ^{-/-} Ig-transgenic B lymphocytes. Cells were incubated with medium alone (lanes U), 1 $\mu\text{g ml}^{-1}$ HEL (H), 10 $\mu\text{g ml}^{-1}$ polyclonal anti-IgM (M), or 50 ng ml⁻¹ phorbol ester (P). ERK2, pp90RSK and EGR1 were analysed in lysates 8, 10 and 60 min after stimulation, respectively. Filled arrows indicate slower-migrating active forms of ERK2 and pp90RSK; open arrows correspond to inactive forms^{24,25}. Numbers under lanes indicate per cent of immunoreactive kinase with mobility corresponding to the active form; numbers to the left are molecular mass markers. **b**, Kinetics of ERK2 and pp90RSK activation in CD45^{+/+} and ^{-/-} B cells stimulated with 1 $\mu\text{g ml}^{-1}$ HEL for the indicated times. Equivalent differences in HEL activation of ERK, RSK and EGR1 were seen in 4 separate experiments and also in cells chronically exposed to soluble HEL *in vivo* (not shown). **c**, Flow cytometry of Indo-1-loaded cells was used to monitor intracellular calcium levels as a function of time in freshly isolated splenic B cells before and after addition (arrows) of soluble HEL (upper panels) or polyclonal goat anti-IgM antibodies (lower panels). Left panels show CD45^{+/+} B cells and right panels CD45^{-/-} B cells. **d**, Summary of data showing the delay between stimulation and time when the peak fraction of cells exhibited elevated calcium (x-axis), and the mode intracellular calcium concentration at this time (y-axis). Filled symbols, CD45^{-/-} cells; open symbols, CD45^{+/+} cells (the latter are distinguished by an oblique slash). Dashed lines represent the mean basal calcium concentration. Data are from three independent experiments; each experiment is indicated by a different symbol. **e**, Splenocytes were incubated in the absence or presence of the indicated concentrations of HEL antigen or in the presence of PdBu (5 ng ml⁻¹) and ionomycin (750 ng ml⁻¹) (P+I) for 17 h (CD86) or 44 h (CD54, FSC). Forward light scatter (FSC) is an indicator of cell size. Cells were stained for CD86 or CD54 and the mode fluorescence measured on HEL-binding, propidium-iodide-negative B cells. Equivalent differences were seen in **c–e**, with cells from chimaeric animals constructed with mixtures of CD45^{-/-} and Ly5⁺-positive CD45^{+/+} cells.

METHODS. For ERK, RSK and EGR1 analysis, splenic B cells were purified at room temperature by magnetically depleting²⁶ non-B cells with FITC-conjugated antibodies to CD4, CD8, Thy1, Mac1 (Caltag Laboratories, South San Francisco) and Gran-1 (Pharmingen, San Diego); and stimulated at 37 °C with HEL or affinity-purified goat anti-IgM (Jackson ImmunoResearch). Cells were lysed in ice-cold hypotonic buffer with 1% NP40 and protease and phosphatase inhibitors. Cytosolic proteins were resolved by SDS-PAGE, transferred to nitrocellulose and immunoblotted with anti-ERK2 (Transduction Labs) or anti-pp90RSK (ref. 25) and nuclear proteins immunoblotted with anti-EGR1 (Santa Cruz Biotechnology). Immunoblots were developed by ECL (Amersham).

ERK2 and RSK autoradiographs were scanned with a Molecular Dynamics Computing Densitometer. Calcium analysis was conducted as previously²⁶. To determine the time to peak fraction of cells with elevated intracellular calcium, integrations of 10-s windows from the time of antigen addition were performed to count total cells in the window and numbers of cells with intracellular calcium above a threshold value set



~40 nM above the basal level. Time to peak fraction of responding cells was the first 10-s window where the ratio of total cells above the threshold was maximal. Flow cytometric analysis was performed on a FACScan with FACS desk software (Beckman Center Shared FACS Facility). Surface marker staining was performed as previously described^{12,26}. GL1 anti-CD86-phycoerythrin was from Pharmingen.

TABLE 2 Accumulation of CD45-deficient B cells within a polyclonal repertoire of CD45⁺ lymphocytes is promoted by binding soluble HEL autoantigen

Recipient:	Bone marrow ($\times 10^{-4}$)§		Spleen ($\times 10^{-5}$)		Lymph node ($\times 10^{-4}$)		Blood (%)	
	Non	HEL	Non	HEL	Non	HEL	Non	HEL
CD45 ^{-/-} Ig-transgenic/non-transgenic mixed chimaeras*								
Non-transgenic B cells†	19.8 (11.0)	20.8 (14.1)	263 (61.1)	206 (26.5)	189 (58.1)	199 (60.0)	35.0 (6.0)	29.2 (5.8)
HEL-binding cells‡	6.5 (3.9)	7.1 (4.1)	26.7 (9.5)	58.6 (20.5)	6.3 (3.0)	39.0 (18.1)	1.25 (0.3)	3.6 (0.6)
			<i>P</i> < 0.01		<i>P</i> < 0.002		<i>P</i> < 0.001	
CD45 ^{-/-} Ig-transgenic/non-transgenic mixed chimaeras*								
Non-transgenic B cells	16.1 (7.8)	18.9 (12.8)	227 (32.0)	263 (93.9)	269 (64.1)	314 (47.7)	27.6 (6.7)	26.1 (6.1)
HEL-binding cells	6.3 (2.9)	8.0 (3.5)	67.5 (33.8)	40.2 (19.9)	36.1 (23.4)	7.9 (2.9)	3.3 (1.2)	1.4 (0.3)

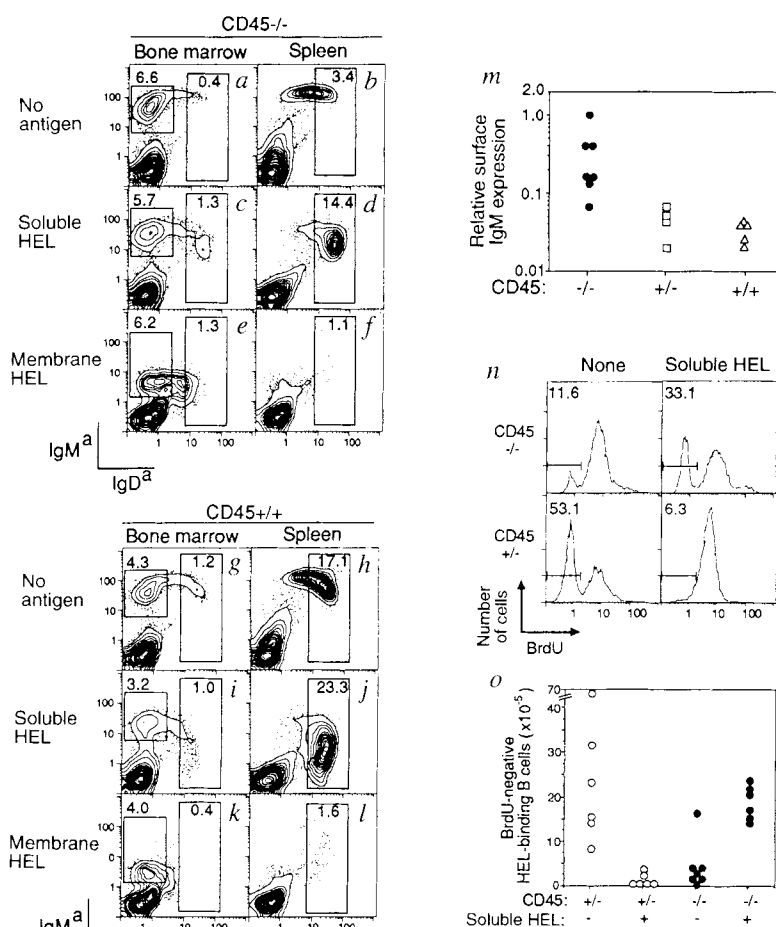
* Haemopoietic radiation chimaeras were generated using an equal mixture of CD45^{-/-} or CD45^{-/-} Ig-transgenic bone marrow and CD45^{+/+} Ly5^a non-transgenic bone marrow to reconstitute lethally irradiated non-transgenic (Non) or soluble-HEL transgenic (HEL) recipients.

† Non-transgenic B cells represent Ly5^a-positive, CR1/2-positive, HEL-negative cells in peripheral tissues and Ly5^a-positive B220-positive, CR1/2-negative cells in bone marrow.

‡ HEL-binding cells represent Ly5^a-negative, HEL-binding CR1/2-positive cells in peripheral tissues and Ly5^a-negative, HEL-binding CR1/2-negative cells in bone marrow.

§ Number of cells in bone marrow from one tibia. Data shown are arithmetic means with standard deviations in parentheses, and are from 6 mice per group. Statistical comparisons are by Student's *t*-test. Data are pooled from two experiments.

FIG. 2 Development of HEL-specific CD45-deficient B cells in the absence or presence of HEL autoantigen. *a-l*, B-cell development in irradiation chimaeras lacking HEL (*a, b, g, h*) or expressing 1 nM soluble HEL autoantigen (*c, d, i, j*) or membrane HEL autoantigen (*e, f, k, l*) 8 weeks after reconstitution with 100% CD45^{-/-} (*a-f*) or CD45^{+/+} (*g-l*) Ig-transgenic bone marrow. Two-colour flow cytometry of cells from bone marrow (left panels) or spleen (right panels) stained for IgM^a and IgD^a. Percentages of immature IgM^a-positive, IgD^a-negative cells in the bone marrow, and of IgD^a-intermediate cells in the spleen and bone marrow, are shown in the windows. In the bone marrow, owing to the large number of immature IgD^a-intermediate B cells, not all the cells gated IgD^a are mature recirculating B cells. The larger fraction of IgD^a-intermediate cells in membrane HEL recipients of CD45^{-/-} bone marrow was a consistent observation in six chimaeric animals and has also been described for cells expressing a *bc/2* transgene²⁷, possibly reflecting a reduced rate of elimination. The small number of IgD^a cells detectable in the spleen of membrane HEL mice were non-HEL-binding variants¹⁴. *m*, IgM modulation in chimaeric mice reconstituted with CD45^{-/-}, ^{-/-} or ^{+/+} bone marrow. Individual points represent the relative level of IgM expressed by cells developing in a soluble HEL transgenic chimaera compared to the level of IgM on B cells from a chimaera reconstituted with the same marrow but lacking HEL autoantigen. *n, o*, Increased numbers of long-lived CD45-deficient B cells in the spleens of autoantigen-expressing recipients. Chimaeric mice constructed with mixtures of Ig-transgenic and non-transgenic bone marrow, as for Table 2, were maintained on water containing BrdU for 7 days before splenocyte isolation and staining for HEL binding and BrdU. In *n*, gates were plotted around HEL-binding cells and BrdU fluorescence was plotted against relative cell number. Cells remaining unlabelled with BrdU after 7 days were gated as indicated by the horizontal line, based upon staining of cells from an animal that did not receive BrdU, and the percentage of unlabelled HEL-binding cells is shown in each panel. The mean percent BrdU-negative cells in the CD45^{-/-} population in the absence of HEL was 16% (*n* = 7) and was 35% (*n* = 6) in the presence of HEL. In *o*, the total number of BrdU-negative HEL-binding cells per spleen was calculated by multiplying the per cent BrdU-negative HEL-binding cells, determined as in *n*, by the total number of splenocytes. Each point represents an individual animal. A similar promotion of CD45^{-/-} cell survival was observed in AL3 mice that express 10 nM soluble HEL (ref. 13), consistent with the weak signalling induced by saturating concentrations of soluble HEL in Fig. 1. METHODS. Ig-transgenic, soluble-HEL (1 nM) transgenic, and membrane-HEL transgenic mice were of the MD4 line¹¹, ML5 line^{11,13} and KLK4 line¹⁴, respectively. Transgenic mice were maintained on a C57BL/6 (B6) background. CD45^{-/-} and CD45^{-/-} Ig-transgenic mice were second or third



generation intercrosses between B6 transgenic mice and B6 × 129 CD45^{-/-} mice. Multiple sets of chimaeras were constructed with bone marrow from 6 independent ^{-/-} and 4 independent ^{+/+} intercross donors to control for variation in background genotype. Moreover, because HEL and non-transgenic recipients received bone marrow from a common donor in each case, the effects of HEL were isolated from any effects of background genotype. BrdU (Sigma) administration and staining has been described⁷, except that here cells were stained first with unconjugated HEL and HyHEL9-biotin.

type Ig-transgenic bone marrow to reconstitute recipients that lacked or expressed HEL as an autoantigen^{13,14}. In the absence of HEL, immature IgM⁺ IgD⁻ B cells were present in the bone marrow in equivalent numbers in CD45^{-/-}, +/- or +/+ animals (Fig. 2a, g and Table 1). However, there were significantly fewer mature IgD^{hi} B cells in the spleens of CD45^{-/-} chimaeras (Fig. 2b, h and Table 1). Strikingly, accumulation of mature IgD^{hi} CD45^{-/-} cells was promoted in mice expressing soluble HEL autoantigen (Fig. 2d and Table 1). Modulation of cell-surface IgM was less marked on CD45^{-/-} B-cells in the presence of circulating HEL (1–10-fold compared with 10–50-fold on controls; Fig. 2c, d, i, j, m), consistent with the evidence for reduced signalling found *in vitro*. In contrast to the rescuing effect of soluble HEL, in transgenic mice expressing the more potent high-valency form of autoantigen, membrane HEL (refs 14 and 15), IgM modulation, developmental arrest and elimination of immature CD45^{-/-} B cells still occurred (Fig. 2e, f, k, l).

In a B-cell repertoire composed of many competing specificities, soluble HEL autoantigen mediates negative selection of newly produced HEL-binding B cells by blocking their migration into peripheral lymphoid follicles and preventing their survival for more than three days¹⁰. Differential survival of B cells at this stage has also been suggested to involve positive selection by antigen^{16–18}. To analyse selection of CD45^{-/-} B cells in a diverse B-cell repertoire, chimaeric mice were constructed with CD45^{-/-} or CD45^{+/-} Ig-transgenic bone marrow mixed with CD45^{+/-} non-transgenic bone marrow to provide a source of competing non-HEL binding B cells. Similar numbers of immature CD45^{-/-} or CD45^{+/-} HEL-binding cells were present in the bone marrow in animals lacking or expressing HEL autoantigen (Table 2). As observed in the 100% chimaeras (Table 1), in the absence of HEL autoantigen, significantly fewer CD45^{-/-} HEL-specific cells were detectable in spleen, lymph node and blood compared to the numbers of CD45^{+/-} cells in control chimaeras (Table 2). In the presence of circulating HEL, CD45^{-/-} cells were reduced in number in the periphery due to negative selection by competition for entry into follicles¹⁰. Strikingly, rather than being decreased, the frequency of CD45^{-/-} Ig-transgenic B cells was increased 2-fold in the spleen and 3–6-fold in lymph nodes and blood in the presence of circulating HEL autoantigen, reaching numbers equivalent to those of CD45^{+/-} Ig-transgenic B cells in the absence of antigen (compare numbers in bold in Table 2). To establish whether CD45-deficient B cells positively selected by autoantigen had an increased lifespan, the number of splenic B cells that survived for seven or more days was determined by measuring the percentage of HEL-binding cells that remained unlabelled during continuous bromodeoxyuridine (BrdU) feeding (Fig. 2n). Whereas the number of CD45^{+/-} HEL-binding cells that survived was markedly reduced when soluble HEL was present, the same autoantigen promoted the survival of CD45^{-/-} B cells (Fig. 2n, o).

Our data support the idea that signalling from the antigen receptor is required for normal accumulation of mature B cells, and that the amount of signalling may determine positive or negative selection. Previous data, as well as our results, indicate that CD45 is a positive regulator of antigen-receptor signalling. We show here that in CD45-deficient cells, engagement of antigen receptors with soluble antigen induces a weak signal that is sufficient to promote the maturation and/or survival of B cells. These results nevertheless allow for the alternative possibility that qualitative differences in signalling determine B-cell negative or positive selection. The fact that CD45^{-/-} HEL-specific B cells undergo this positive-selection step in the complete absence of HEL may mean that their receptors signal weakly but spontaneously or that, like positive selection of T cells^{19–21}, the HEL-specific receptor crossreacts weakly with one or more endogenous autoantigens. Interestingly, in CD45-deficient non-transgenic mice, mature B-cell numbers are not reduced¹. Furthermore, in mixed chimaeras constructed with non-transgenic CD45^{-/-} and CD45^{+/-} bone marrow, 65% of the CD45-deficient B cells remain

unlabelled with BrdU after seven days. The possibility that a repertoire of mature B cells is selected in these animals which has increased affinity for self-antigens is being investigated. Mutations or genetic polymorphisms that reduce antigen-receptor signal strength, exemplified by the CD45 mutation here or a mutation in *lyn* tyrosine kinase which results in autoantibody production^{22,23}, may constitute an important predisposing factor to autoimmune disease by favouring retention of higher-affinity self-reactive B cells in the recirculating repertoire. □

Received 3 November 1995; accepted 4 April 1996.

1. Kishihara, K. *et al.* *Cell* **74**, 143–156 (1993).
2. Pingel, J. T. & Thomas, M. L. *Cell* **58**, 1055–1065 (1989).
3. Koretzky, G. A., Picus, J., Thomas, M. L. & Weiss, A. *Nature* **346**, 66–68 (1990).
4. Peyron, J.-F. *et al.* *Int. Immunol.* **3**, 1357–1366 (1991).
5. Shiroy, M., Lindsey, G., Biffen, M., Shivan, E. & Alexander, D. *EMBO J.* **11**, 4887–4897 (1992).
6. Volarevic, S. *et al.* *J. exp. Med.* **176**, 835–844 (1992).
7. Justement, L. B., Campbell, K. S., Chien, N. C. & Cambier, J. C. *Science* **252**, 1839–1842 (1991).
8. Ogimoto, M. *et al.* *Eur. J. Immun.* **25**, 2265–2271 (1995).
9. Chui, D., Ong, C. J., Johnson, P., Teh, H.-S. & Marth, J. D. *EMBO J.* **13**, 798–807 (1994).
10. Cyster, J. G., Hartley, S. B. & Goodnow, C. C. *Nature* **371**, 389–395 (1994).
11. Goodnow, C. C. *et al.* *Nature* **334**, 676–682 (1988).
12. Gold, M. R. & DeFranco, A. L. *Adv. Immun.* **55**, 221–295 (1994).
13. Adelstein, S. *et al.* *Science* **251**, 1223–1225 (1991).
14. Hartley, S. B. *et al.* *Nature* **353**, 765–769 (1991).
15. Hartley, S. B. & Goodnow, C. C. *Int. Immunol.* **6**, 1417–1425 (1994).
16. MacLennan, I. C. M. & Gray, D. *Immunol. Rev.* **91**, 61–85 (1986).
17. Malyon, B. A., Yancopoulos, G. D., Barth, J. E., Bona, C. A. & Alt, F. W. *J. exp. Med.* **171**, 843–859 (1990).
18. Gu, H., Tarlinton, D., Muller, W., Rajewsky, K. & Forster, I. *J. exp. Med.* **173**, 1357–1371 (1991).
19. Hogquist, K. A. *et al.* *Cell* **76**, 17–27 (1994).
20. Ashton-Rickardt, P. G. *et al.* *Cell* **76**, 651–663 (1994).
21. Sebzdka, E. *et al.* *Science* **263**, 1615–1618 (1994).
22. Hibbs, M. L. *et al.* *Cell* **83**, 301–311 (1995).
23. Nishizumi, H. *et al.* *Immunity* **3**, 549–560 (1995).
24. Posada, J. & Cooper, J. *Science* **225**, 212–215 (1992).
25. Chen, R., Chung, J. & Bilenis, J. *Molec. cell. Biol.* **11**, 1861–1867 (1991).
26. Cooke, M. P. *et al.* *J. exp. Med.* **179**, 425–438 (1994).
27. Hartley, S. B. *et al.* *Cell* **72**, 325–335 (1993).

ACKNOWLEDGEMENTS. We thank L. White for genotyping and maintenance of mice, J. Bilenis for antiserum to pp90RSK, M. Davis, R. Glynn, N. Killeen and C. Somoza for comments on the manuscript, and the divisions of laboratory animal medicine at Washington University and Stanford for animal husbandry. J.G.C. was supported by a Cancer Research Institute postdoctoral fellowship. J.I.H. is supported by an MSTP training grant funded by the National Institute for General Medical Sciences. M.L.T. is supported by a public health grant. C.C.G. and M.L.T. are investigators of the Howard Hughes Medical Institute.

CORRESPONDENCE and requests for materials should be addressed to C.C.G. (e-mail: hf.exr@forsythe.stanford.edu). The ML4, MD5 and KLR4 transgenic lines are available from the Induced Mutant Repository of the Jackson Laboratories.

Effects on NF- κ B1/p105 processing of the interaction between the HTLV-1 transactivator Tax and the proteasome

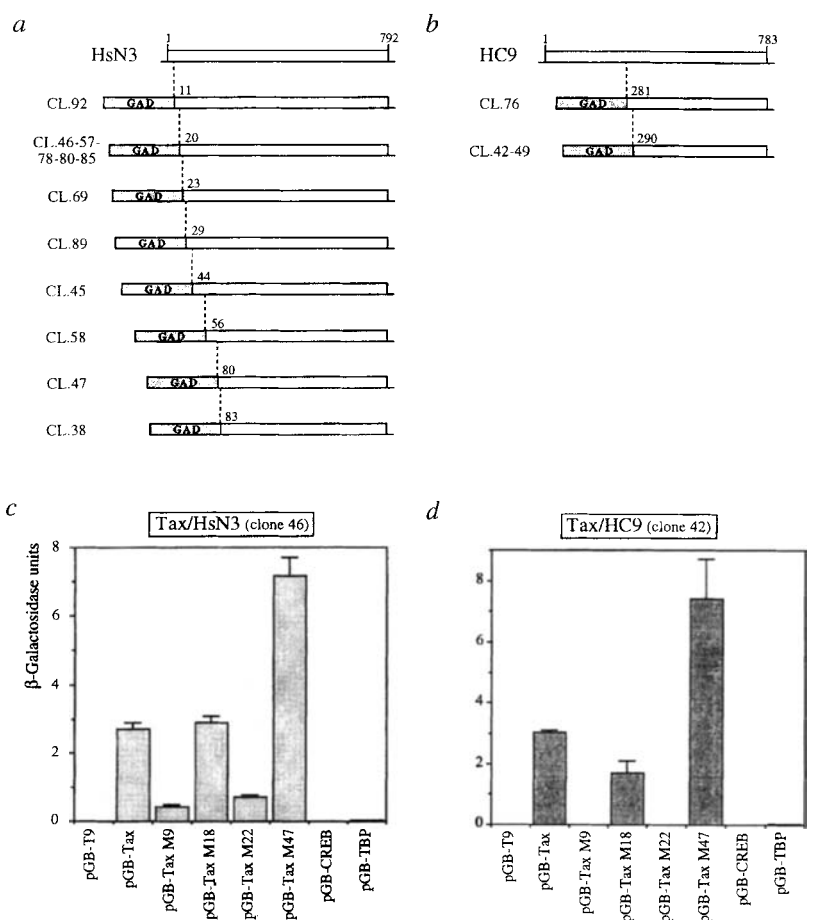
Raphaël Rousset, Christelle Desbois, Frédéric Bantignies & Pierre Jalinot

CNRS UMR49, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

THE viral Tax protein, which is encoded by human T-cell leukaemia virus HTLV-I, activates nuclear translocation of the NF- κ B/Rel transcription factors and relieves cytoplasmic sequestration of RelA and Rel by heterodimerization with NF- κ B1/p105 (refs 1,2). Proteolytic maturation of this precursor protein is performed by the proteasome complex^{3,4}. Here we show that Tax binds specifically to two subunits of the 20S proteasome, HsN3 and HC9. This interaction is weakened with HsN3 and lost for HC9 when a mutant of Tax is substituted that is selectively

FIG. 1 Isolation by yeast two-hybrid screening of two subunits of the 20S proteasome, HsN3 and HC9, that interact with Tax. *a*, *b*, Clones isolated corresponding to HsN3 (*a*) and to HC9 (*b*). For HsN3, eight different cDNAs were obtained, one of which was found five times. Of three HC9 cDNAs, two were identical. The numbers shown above each clone indicate the first nucleotides of the HsN3 and HC9 cDNAs with respect to the first nucleotide of the translation initiation codon. In all these clones, the HsN3 and HC9 coding sequences were in the same reading frame as that of the GAL4 transcriptional activation domain (GAD). *c*, Analysis of the interaction between HsN3 and various point mutants of Tax using a liquid culture β -galactosidase assay. Yeasts were transformed by the HsN3 clone CL.46 plus plasmids expressing the GAL4 DNA-binding domain (GB) either alone (pGB-T9) or fused to Tax. The Tax sequence was either wild type (pGB-Tax) or included the following mutations: M9 (⁴¹His-Arg changed to Ala-Ser: pGB-Tax M9), M18 (¹²³Thr-Leu changed to Ala-Ser: pGB-Tax M18), M22 (¹³⁷Gly-Leu changed to Ala-Ser: pGB-Tax M22), and M47 (³¹⁹Leu-Leu changed to Arg-Ser: pGB-Tax M47)⁸. The HsN3 clone was also co-transformed with plasmids expressing fusion proteins associating GB with the rat cAMP-responsive element binding protein (pGB-CREB) and with the human TATA-box-binding protein (pGB-TBP). Expression of β -galactosidase in these various transformants is given in Miller units²⁰. Measurements were made on two independent transformants for each combination of plasmids; values are means $\pm$ s.e. *d*, As *c*, but with the HC9 clone CL.42.

METHODS. *a*, *b*, The Tax coding sequence was inserted in the yeast expression vector pGB-T9 (Clontech). Construction of the library of cDNAs from EBV-transformed human peripheral lymphocytes in λ -ACT has been described²¹. Transformation of yeast strain HF7c (Clontech) by pGB-Tax and the library of cDNAs was performed using a variation of the lithium acetate method²². The HF7c strain includes the *HIS3* and *lacZ* genes under the control of three GAL4-binding sites. Transformants were first selected on a minimal medium (SD) lacking histidine. Out of 250,000 estimate transformants, about 700 colonies were His⁺-positive. These colonies were then assayed on filters for β -galactosidase activity²³ and 94 colonies His⁺ blue were isolated. pACT plasmids were extracted and sequenced. *c*, *d*, Sequences of the Tax mutants M9, M18, M22 and M47, and of CREB and TBP were inserted into the pGB-T9 vector in fusion with GB. Co-



transformations were performed in yeast strain SFY526 (Clontech), which includes the *lacZ* reporter gene under the control of the *GAL1* promoter. β -Galactosidase activity was assayed in liquid culture using *O*-nitrophenyl- β -D-galactoside as substrate²⁴.

FIG. 2 Tax-mediated stimulation of p105 processing in mammalian cells. COS-7 cells were transfected with a p105 expression construct minus or plus a Tax expression vector. Vectors expressing Tax wild type and including mutations M22 or M47 (ref. 8) were used. Cells were pulse-labelled with a [³⁵S]methionine-cysteine mix and collected either immediately or after a 1- or 4-h chase with cold methionine and cysteine. Cells were lysed in RIPA buffer²⁵ and lysates immunoprecipitated using an antibody directed against p105 (ref. 26) which also reacts with p50. Immunoprecipitated proteins were separated by SDS-PAGE. *a*, *b*, Images after exposure of gels to a phosphor screen. *c*, Signals corresponding to p50 and p105 were quantified and their ratio plotted against time. **METHODS.** The pSG-p105 and pSG-Tax²⁷ expression vectors are derivatives of pSG5 (ref. 28). COS-7 cells were cultured in 100-mm plates and transfected using calcium phosphate coprecipitation²⁷. The DNA mix included 2 μ g pSG-p105 plus 500 ng of one of the following plasmids: pSG5, pSG-Tax, pSG-Tax M22, pSG-Tax M47. After removal of the precipitate, cells were cultured in medium complemented with 5% FCS. Pulse labelling was performed by using 0.2 mCi PRO-MIX (Amersham) in the culture medium for 20 min. Cells were lysed (*t* = 0) or replenished for either 1 or 4 h with medium containing cold excess methionine and cysteine. Immunoprecipitation in RIPA buffer using anti-p105 antibody and protein A-Sepharose beads has been described²⁵. Proteins were resolved on 8% polyacrylamide gels and band intensities quantified with a Storm 840 phosphor-Imager (Molecular Dynamics).

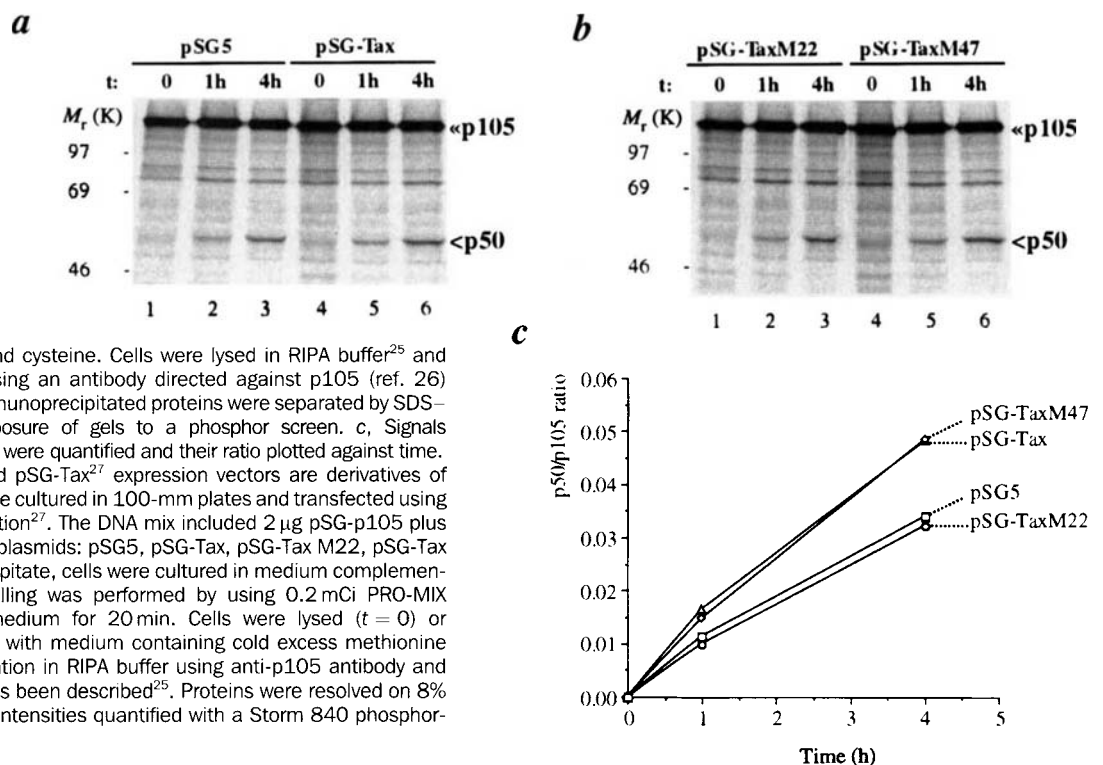
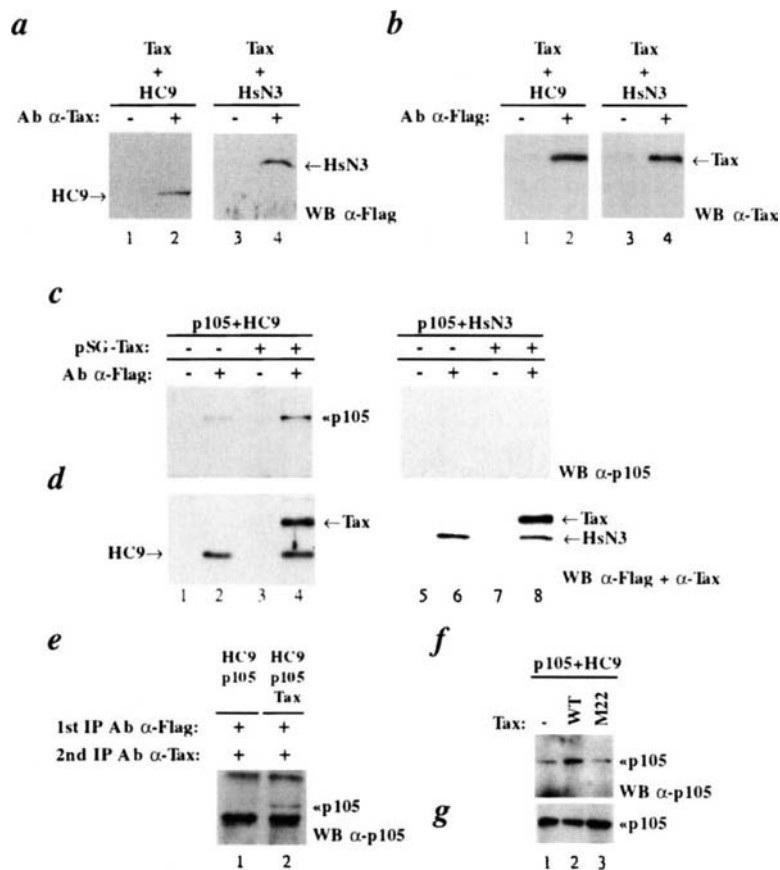


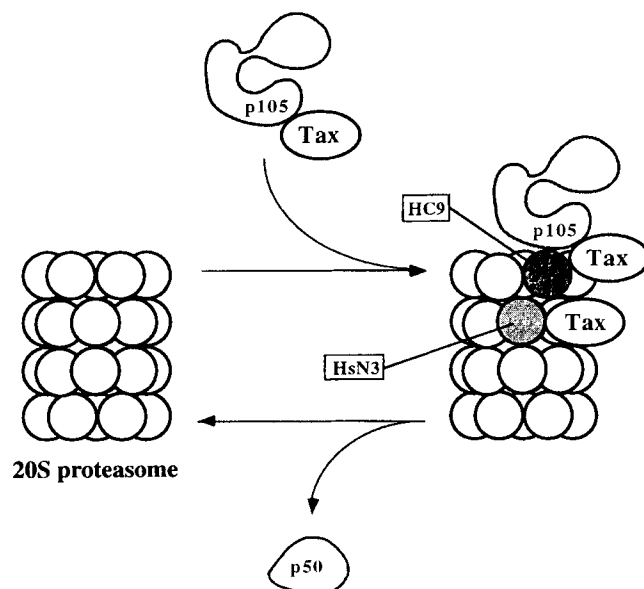
FIG. 3 Tax increased p105 binding to the HC9 subunit of the 20S proteasome. *a, b*, Interaction of HC9 (lanes 1 and 2) and HsN3 (lanes 3 and 4) with Tax in COS-7 cells. Cells were transfected with pSG-Tax and a vector expressing either HsN3 (pSGF-HsN3) or HC9 (pSGF-HC9) tagged with the Flag epitope at the N terminus. In *a*, immunoprecipitation was done in RIPA buffer²⁵ using an anti-Tax antibody²⁹ (Ab α -Tax) and analysis of the immunoprecipitated (IP) proteins resolved in a 10% polyacrylamide-SDS gel was by western blotting with the monoclonal antibody M2 (Kodak) directed against the Flag epitope (WB α -Flag). As controls, the anti-Tax antibody was omitted (odd-number lanes). In *b*, immunoprecipitations were done with the anti-Flag antibody (Ab α -Flag) and western blotted with antiserum against Tax (WB α -Tax). *c*, Analysis of the association of p105 with either HC9 or HsN3 in the presence and the absence of Tax. COS-7 cells were transfected with either pSGF-HC9 (lanes 1 to 4) or pSGF-HsN3 (lanes 5 to 8) and pSG-p105. pSG-Tax was omitted or added to the plasmid mix as indicated. Immunoprecipitations were done in RIPA buffer²⁵ using the anti-Flag monoclonal antibody (even-number lanes). As controls, the anti-Flag antibody was omitted (odd-number lanes). The immunoprecipitated proteins were separated by SDS-PAGE and transferred to a nitrocellulose membrane which was successively probed with antibody directed against p105 (*c*, WB α -p105) and a mix of the antibodies directed against Tax and Flag (*d*, WB α -Flag + α -Tax). Positions of p105, Tax and either HC9 or HsN3 are indicated. *e*, Analysis of the association between Tax and p105 in complex with HC9. Cells were transfected with pSGF-HC9 and pSG-p105 minus (lane 1) or plus pSG-Tax (lane 2). Immunoprecipitation was in RIPA buffer using anti-Flag antibody; immunoprecipitated complexes were eluted by incubation with an excess of the Flag peptide. A second immunoprecipitation was done with antibody directed against Tax. Immunoprecipitated proteins were analysed by western blotting using the antibody against p105. The band below the p105 signal is nonspecific and is due to a reaction of anti-p105 antibody against anti-Tax. *f, g*, Effect of the M22 mutation on the stimulation by Tax of the binding of p105 to HC9. The experiment shown in *c* was done with HC9 and Tax which was either wild type (lane 2) or included mutation M22 (lane 3). Western blots with antibody against p105 of the IP complexes (*f*) and of crude lysates (*g*) are shown. In these lysates, mutant Tax was expressed at the same level as wild type, as evaluated by western blotting with anti-Tax antibody (data not shown).

METHODS. To isolate full-length HC9 cDNA, 2 μ g total RNA from Jurkat-T cells was reverse-transcribed with a specific HC9 antisense oligonucleotide.



HC9 coding sequence was then amplified by PCR and cloned in the pSG5 derivative pSG-Flag, which includes Flag epitope (Kodak), giving pSGF-HC9. Amplification of HC9 cDNA was verified by sequencing. pSGF-HsN3 was obtained by subcloning the cDNA insert of CL46 in pSG-Flag. COS-7 cells were transfected with 1 μ g of each expression vector. Elution of complexes precipitated by anti-Flag antibody was done by incubation for 2 h with a 500-fold molar excess of Flag peptide (with respect to antibody). Western blots were prepared using the ECL system (Amersham).

FIG. 4 Model for the mechanism of action of Tax on p105 processing by the proteasome. The 28 subunits of the 20S proteasome are arranged into 4 heptameric rings^{16,17}. The top and bottom outer rings are composed of 7 α -subunits and the two inner rings contain 7 β -subunits¹⁸. Based on sequence comparisons, HC9 and HsN3 have been classified as α - and β -subunits, respectively^{6,7}. HsN3 in the inner rings has been precisely localised¹⁹. On the basis of these data and our results we propose that the dual interaction of Tax with p105 and HC9 favours recruitment of p105 to the proteasome as shown here, which would increase processing of p105 into p50. As Tax can form homodimers³⁰, one Tax protein may interact with HC9 and another with p105 (not shown).



defective for NF- κ B activation. Immunoprecipitation shows that p105 binds weakly to HC9 and that this interaction is reinforced by Tax. No bridging function of Tax between p105 and HsN3 was observed. From these results, we propose that Tax accelerates the proteolytic maturation of p105 by favouring its anchorage to the proteasome.

To find cellular proteins that interact with Tax, we used yeast two-hybrid screening⁵ with a fusion protein that combined the GAL4 DNA-binding domain with Tax. Several positive clones were found to correspond to two subunits of the 20S proteasome, namely HsN3 (ref. 6) and HC9 (ref. 7) (Fig. 1*a, b*). For both HsN3 and HC9, there was no interaction with fusion proteins that contained the GAL4 DNA-binding domain fused to either CREB or TBP, indicating that the interaction between Tax and these two proteasome subunits is specific (Fig. 1*c, d*).

To understand the significance of the interaction between Tax and these proteins, various point mutants of Tax⁸ were tested. Tax activates various types of *cis*-regulatory promoter elements, including CRE, SRE and κ B sites⁹. Mutation M22 specifically impairs the effect of Tax on NF- κ B-dependent transcription relative to its activity on ATF/CREB factors⁸. Conversely, Tax M47 is unable to activate the HTLV-I promoter but is active on the NF- κ B pathway⁸. Tax M9 and M18 do not transactivate either the HTLV-I or HIV-1 promoters⁸. However, Tax M18 is about half as active as wild type when tested on the κ B site of the interleukin-2 receptor- α gene (data not shown). Using the two-hybrid system, mutations M9 and M22 were found to weaken the interaction between Tax and HsN3, whereas mutations M18 and M47 did not (Fig. 1*c*), M47 even having a positive effect. Mutations M9 and M22 abolished the interaction with HC9 (Fig. 1*d*); M18 reduced it by half. As with HsN3, M47 strengthened the interaction. Therefore a correlation exists between the ability of Tax to interact with HC9 and the effect of this transactivator on the activation of the NF- κ B pathway. This correlation is less evident for HsN3.

Tax induces nuclear translocation of NF- κ B factors^{10–12}. As heterodimers with NF- κ B1/p105, RelA and Rel proteins are sequestered in the cytoplasm^{13,14}. Tax relieves this negative effect of p105 and allows nuclear translocation of Rel and RelA (refs 1, 2); it also binds p105 (ref. 15). We therefore investigated whether the interaction of Tax with the proteasome subunits plays a part in the reversal by Tax of the p105-mediated inactivation of Rel proteins. We first tested whether Tax can increase p105 processing. COS-7 cells were transfected with a vector expressing p105 and were pulse-labelled using ³⁵S-methionine and ³⁵S-cysteine cell extracts were immunoprecipitated with an antibody

against p105. The results show that coexpression of Tax does increase the amount of p105 processed into p50 (Fig. 2). Tax M47 had the same activity as wild type, but mutation M22 completely abolished it (Fig. 2*c*).

We next tested whether this interaction in yeast of Tax and the proteasome subunits occurred in mammalian cells. COS-7 cells were co-transfected with vectors expressing Tax and either HsN3 or HC9 fused to a Flag epitope. HsN3 and HC9 were found to immunoprecipitate with Tax (Fig. 3*a, b*). This binding was also observed *in vitro* using Tax produced in bacteria as a fusion protein with glutathione *S*-transferase (data not shown). As Tax interacts with p105, we investigated whether it could bridge p105 and the proteasome subunits. The vector expressing either HC9 or HsN3 fused to the Flag epitope was co-transfected with those producing Tax and p105. Proteins were immunoprecipitated using an anti-Flag antibody and analysed by western blotting. In the absence of Tax, HC9 bound weakly to p105 (Fig. 3*c*, lane 2), but binding was stronger in the presence of Tax (Fig. 3*c*, lane 4). Sequential precipitation of HC9 and Tax showed that Tax and p105 are associated in the ternary complex (Fig. 3*e*, lane 2). Tax M22 was unable to stimulate formation of this ternary complex (Fig. 3*f*, lanes 2 and 3). The amount of p105 detected in the crude lysates was not modified by the presence of Tax (Fig. 3*g*). These results indicate that p105 weakly binds HC9 and that this interaction is strengthened by Tax. In contrast, there was no association between p105 and HsN3, either in the presence or absence of Tax (Fig. 3*c, d*, lanes 6 and 8).

Our results show that the viral transactivator Tax interacts with two subunits of the 20S proteasome. This multicatalytic complex is a cylinder composed of 28 subunits arranged into 4 heptameric rings^{16,17}. The α -subunit HC9 is located in the outer ring of this structure; the β -subunit HsN3 is in the inner ring^{18,19} (Fig. 4). *In vivo*, Tax strengthens a weak interaction between p105 and HC9 and stimulates the processing of p105 to p50, supporting the idea that Tax helps anchor p105 to the proteasome by interacting with both p105 and HC9 (Fig. 4). The function of the interaction between Tax and HsN3 in the processing of p105 is not known, but it may contribute by modifying the activity of the proteasome. As p105 has to be ubiquitinated before it can be processed by the 26S proteasome^{3,4}, Tax may induce a direct loading of p105 onto the 20S proteasome or favour binding of the ubiquitinated form to the 26S proteasome. In conclusion, Tax represents the first example of a factor regulating maturation of a precursor protein by the proteasome. Analysis of the interaction between p105 and HC9 should enable the function of the proteasome in this processing to be clarified. □

Received 6 November 1995; accepted 4 April 1996.

- Watanabe, M. et al. *Oncogene* **8**, 2949–2958 (1993).
- Muñoz, E., Courtois, G., Veschambre, P., Jalilot, P. & Israël, A. *J. Virol.* **68**, 8035–8044 (1994).
- Palombella, V. J., Rando, O. J., Goldberg, A. L. & Maniatis, T. *Cell* **78**, 773–785 (1994).
- Orian, A. et al. *J. Biol. Chem.* **270**, 21707–21714 (1995).
- Fields, S. & Song, O.-K. *Nature* **340**, 245–246 (1989).
- Nothwang, H. G., Tamura, T., Tanaka, K. & Ichihara, A. *Biochim. biophys. Acta* **1219**, 361–368 (1994).
- Tamura, T. et al. *Biochim. biophys. Acta* **1089**, 95–102 (1991).
- Smith, M. R. & Greene, W. C. *Genes Dev.* **4**, 1875–1885 (1990).
- Yoshida, M., Suzuki, T., Fujisawa, J. & Hirai, H. *Curr. Top. Microbiol. Immun.* **193**, 79–89 (1995).
- Ballard, D. W. et al. *Science* **241**, 1652–1655 (1988).
- Leung, K. & Nabel, G. J. *Nature* **333**, 776–778 (1988).
- Ruben, S. et al. *Science* **241**, 89–91 (1988).
- Rice, N. R., MacKichan, M. L. & Israël, A. *Cell* **71**, 243–253 (1992).
- Mercurio, F., DiDonato, J., Rosette, C. & Karin, M. *Genes Dev.* **7**, 705–718 (1993).
- Hirai, H. et al. *Oncogene* **7**, 1737–1742 (1992).
- Pühler, G. et al. *EMBO J.* **11**, 1607–1616 (1992).
- Löwe, J. et al. *Science* **268**, 533–539 (1995).
- Grziwa, A., Baumeister, W., Dahlmann, B. & Kopp, F. *FEBS Lett.* **290**, 186–190 (1991).

- Kopp, F. et al. *J. molec. Biol.* **248**, 264–272 (1995).
- Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, NY 1972).
- Durfee, T. et al. *Genes Dev.* **7**, 555–569 (1993).
- Gietz, D., St Jean, A., Woods, R. A. & Schiestl, R. H. *Nucleic Acids Res.* **20**, 1425 (1992).
- Breedon, L. & Nasmyth, K. *Cold Spring Harb. Symp. quant. Biol.* **50**, 643–650 (1985).
- Guarente, L. *Meth. Enzym.* **101**, 181–191 (1983).
- Harlow, E. & Lane, D. in *Antibodies: A Laboratory Manual* 447–470 (Cold Spring Harbor Laboratory Press, NY, 1988).
- Kieran, M. et al. *Cell* **62**, 1007–1018 (1990).
- Crenon, I. et al. *Oncogene* **8**, 867–875 (1993).
- Green, S., Isseemann, I. & Sheer, E. *Nucleic Acids Res.* **16**, 369 (1988).
- Harly, S. M. et al. *Genes Dev.* **3**, 1534–1544 (1989).
- Perini, G., Wagner, S. & Green, M. R. *Nature* **376**, 602–605 (1995).

ACKNOWLEDGEMENTS. We thank S. Elledge for the cDNA library in pACT; B. Cullen for antiserum against Tax; W.-C. Greene for Tax point mutants; A. Israël for antibody against p105; E. Gilson for introducing us to the manipulation of yeast; and B. B. Rudkin for critical reading of the manuscript. This work was supported by the Agence Nationale de Recherches sur le SIDA and the Association pour la Recherche contre le Cancer (F.B.).

CORRESPONDENCE and requests for materials should be addressed to P.J. (e-mail: pjalilnot@ens-lyon.fr).

Efficient integration of an intron RNA into double-stranded DNA by reverse splicing

Jian Yang*, Steven Zimmerly*, Philip S. Perlman† & Alan M. Lambowitz*

* Departments of Molecular Genetics, Biochemistry, and Medical Biochemistry, The Ohio State University, 484 West Twelfth Avenue, Columbus, Ohio 43210-1292, USA

† Department of Biochemistry, University of Texas, Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75235-9038, USA

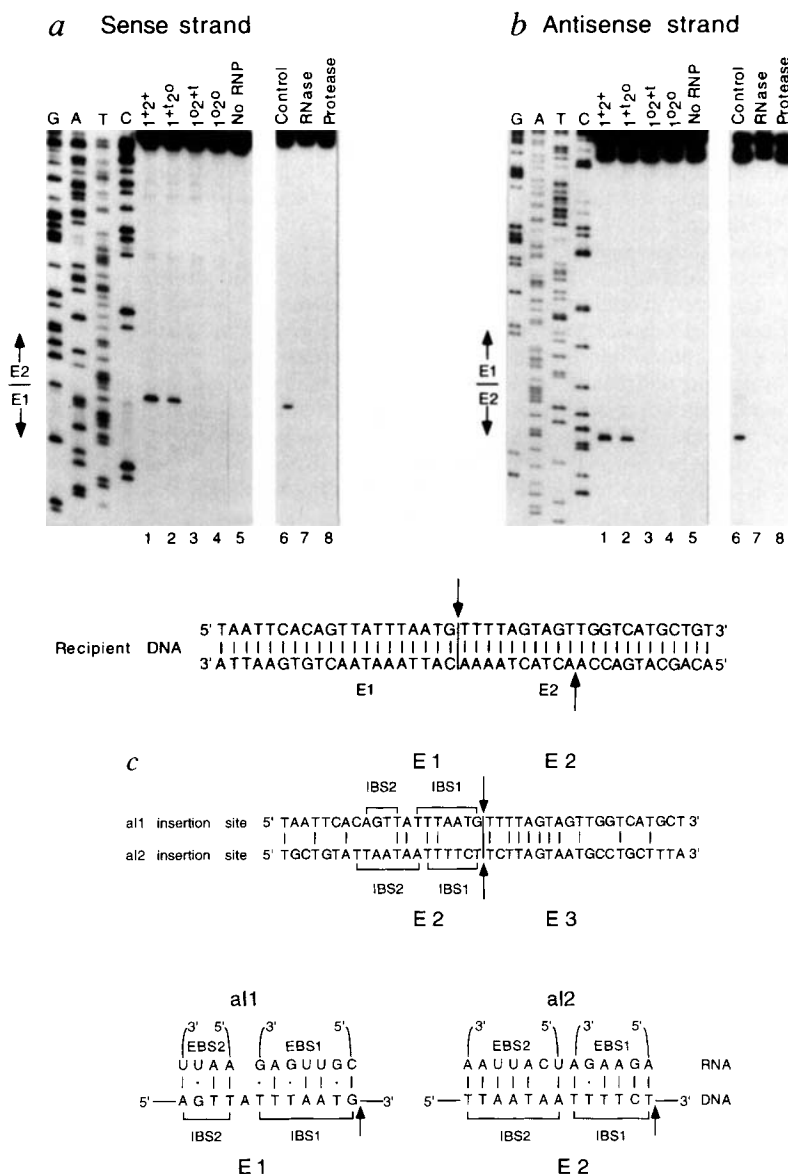
SOME group II introns are mobile elements as well as catalytic RNAs^{1,2}. Introns a11 and a12 found in the gene *COX1* in yeast mitochondria encode reverse transcriptases which promote site-specific insertion of the intron into intronless alleles ('homing')³⁻⁶. For a12 this predominantly occurs by reverse transcription of unspliced precursor RNA at a break in double-strand DNA made by an endonuclease encoded by the intron⁷. The a12 endonuclease involves both the excised intron RNA, which

cleaves the DNA's sense strand by partial reverse splicing; and the intron-encoded reverse transcriptase which cleaves the anti-sense strand⁸. Here we show that a11 encodes an analogous endonuclease specific for a different target site compatible with the different exon-binding sequences of the intron RNA. Over half of a11 undergoes complete reverse splicing *in vitro*, thus integrating linear intron RNA directly into the DNA. This unprecedented reaction has implications for both intron mobility and evolution, and potential genetic engineering applications.

The a12 endonuclease described previously was assayed by cleavage of double-stranded DNA substrates containing the a12 homing site^{7,8}. The RNA-catalysed sense-strand cleavage occurred precisely at the intron insertion site, whereas the protein-catalysed antisense-strand cleavage occurred at position +10 of exon 3. The a11 allele in *Saccharomyces cerevisiae* wild-type strain 161 (1⁺2⁺; for nomenclature, see Fig. 1 legend) is not independently mobile in crosses to a standard 1⁺2⁺ recipient strain with *COX1* exons from *S. capensis*, but is mobile in crosses to a 1⁺2⁺ strain with *COX1* exons from strain 161 (refs 5,6; C1036Δ1Δ2; R. Eskes *et al.*, manuscript in preparation). Thus, to assay the a11 endonuclease activity, we used DNA substrates containing the a11 homing site (E1-E2 junction) from C1036Δ1Δ2. As shown in Fig. 1, a site-specific endonuclease activity that cleaves each strand was detected with 1⁺2⁺ or 1⁺2[°]

FIG. 1 Assay of a11-encoded DNA endonuclease activity. *a*, *b*, Sense- and antisense-strand cleavage, respectively. The DNA substrate was generated by PCR with 5' end-labelled primers to label each strand separately. Lanes: 1-4, reactions with mtRNP particles from strains 161 (1⁺2⁺), 1⁺1⁺2[°], 1⁺2⁺1⁺, C1036Δ1Δ2 (1⁺2[°]); 5, substrate incubated without mtRNP particles; 6-8, 1⁺2⁺ mtRNP particles preincubated without (control) or with RNase or protease. Strains are denoted by a convention in which + indicates the wild-type intron, and ° indicates the absence of the intron; other superscripts indicate specific alleles. Below is the sequence of the a11 homing site with cleavage sites of the sense (top) and antisense (bottom) strands indicated by arrows. *c*, Comparison of the a11 and a12 insertion sites and EBS(RNA)/IBS(DNA) interactions for the two introns.

METHODS. MtRNP particles (0.025 absorbance units at 260 nm) were incubated with 5' ³²P-labelled DNA substrates (300,000 c.p.m., 250 fmol) in 10 μl 50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 50 mM KCl, 5 mM dithiothreitol, for 20 min at 37 °C, as described for the a12 endonuclease⁷. Cleavage products were analysed in denaturing 6% polyacrylamide gels. DNA substrates were synthesized from pE1E2 by PCR and gel-purified as described⁸. pE1E2 contains the E1-E2 junction (69 bp of E1 and 36 bp of E2) amplified from C1036Δ1Δ2 (refs 4,6) mtDNA by PCR with primers E1JY (5'-TAATCATTAGATTAGAATTAGTGCACCTG) and E2JY (5'-AGAAAATCATTATACAGC) and cloned in the *Sma*I site of pKS+ (Stratagene). The sense-strand substrate (152 bp) was synthesized from pE1E2 by PCR with 5'-end-labelled E1JY and unlabelled KS (5'-TCGAGGTCGACGGTATC), and the antisense-strand substrate (176 bp) was synthesized with 5'-end-labelled KS and unlabelled SK (5'-CGCTCTAGAAC-TAGTGGATC). Sequencing ladders were generated from pE1E2 with the corresponding 5' end-labelled primer. For RNase and protease digestions, mtRNP particles were preincubated (2 min, 37 °C) with RNase A (0.2 μg μl⁻¹; Sigma) plus RNase T₁ (0.25 units μl⁻¹; Boehringer) or proteinase K (0.2 μg μl⁻¹; Sigma).



mitochondrial (mt)RNP particles, which contain aI1, but not with $1^{\circ}2^{+}$ or $1^{\circ}2^{\circ}$ mtRNP particles, which lack aI1. The sense strand was cleaved precisely at the intron insertion site, whereas the antisense strand was cleaved at position +10 of the 3' exon, the same relative positions found for the aI2 cleavages⁷. Like the aI2 endonuclease, aI1 activity was abolished by pretreatment with RNase or protease (Fig. 1a, b). The involvement of aI1 protein and a requirement for spliced aI1 RNA were indicated by deficiency of the activity in aI1 mutant strains ($1^{\text{YAHH}2^{\circ}}$, $1^{\text{C}2126}2^{+}$ and $1^{\text{C}1036}2^{\text{C}1036}$; results not shown; see refs 7 and 9, and R. Eskes *et al.* (in preparation) for strain descriptions). In the case of aI2, the 3' end of the cleaved antisense strand is used as a primer for reverse transcription beginning in exon 3 of unspliced precursor RNA. Integration of the resulting cDNA leads to co-conversion of aI2 with flanking exon sequences⁷. For aI1, biochemical assays confirmed that the 3' end of the cleaved antisense strand is also used for target DNA-primed reverse transcription by the aI1-encoded reverse transcriptase (J.Y. *et al.*, manuscript in preparation).

To determine whether the sense-strand cleavage at the aI1 intron insertion site occurs through a reverse splicing reaction, the products were analysed in agarose gels. As shown in Fig. 2a, $1^{\circ}2^{+}$ and $1^{\circ}2^{\circ}$ mtRNP particles gave three ^{32}P -labelled bands of ~2.5 kilobases (kb) (the top band is a closely spaced doublet), the size expected for the intron RNA linked to the small, labelled DNA substrate by partial or complete reverse splicing. The products were not detected with $1^{\circ}2^{+}$ or $1^{\circ}2^{\circ}$ mtRNP particles, which lack aI1 (Fig. 2a), and they have the expected nuclease sensitivities (Fig. 2b). Oligonucleotide-directed digestion with RNase H confirmed that all three products contain aI1 RNA (results not shown). As expected, the ribonucleoprotein (RNP)-mediated reverse-splicing reaction of aI1 was abolished by pre-treating the RNP particles with RNase or protease (Fig. 2c) and was inhibited by the same aI1 mutations that inhibited sense-strand cleavage ($1^{\text{YAHH}2^{\circ}}$, $1^{\text{C}2126}2^{+}$ and $1^{\text{C}1036}2^{\text{C}1036}$; results not shown).

In the case of the aI2 reverse-splicing reaction, agarose gels resolved only two major products, and those appeared to contain two different forms of aI2 lariat linked to the 3' exon DNA via partial reverse splicing⁸. To characterize the aI1 products, reverse-splicing reactions were carried out using DNA substrates individually labelled at each of the four termini. All three products were detected with the 3' sense-strand-labelled substrate, indicating that they contain 3' exon DNA, but only the top band was also detected with the 5' sense-strand-labelled substrate (Fig. 2d, and data not shown). Polyacrylamide gels confirmed that two products containing only 3' exon DNA migrated in the lariat range (Fig. 3a). These presumably correspond to partially reverse-spliced products containing two different forms of intron lariat linked to the 3' exon DNA. The remaining product, which contains both 5' and 3' exon DNA, migrated just above linear intron RNA and had the characteristics expected for aI1 RNA inserted between the DNA exons by complete reverse splicing. Analysis by reverse transcription in conjunction with polymerase chain reaction (RT-PCR) confirmed the expected 5' and 3' intron-exon junction sequences (Fig. 3b–d). In three experiments, the fully reverse-spliced product constituted 38–63% of the total. Although group I and group II introns have been shown to cleave or partially reverse-splice into DNA^{10–12}, our demonstration of complete reverse splicing into DNA is unprecedented for any intron RNA.

To determine whether the fully reverse-spliced product was cleaved on the antisense strand, 5' sense-strand-labelled double-stranded DNAs containing the integrated intron were digested with RNase and analysed in a non-denaturing polyacrylamide gel to detect 'half' molecules resulting from breaks in both strands. The use of 5' sense-strand-labelled DNAs was necessary to distinguish the fully reverse-spliced product from a background of partially reverse-spliced products. Figure 4 shows that > 95% of the fully reverse-spliced product had been cleaved on the antisense strand, as judged by the appearance of correctly sized 5'-half

molecules. The 3' end of the cleaved antisense strand (3' exon position +10; Fig. 1) is positioned to prime reverse transcription of the inserted intron RNA, and we confirmed that the partially or fully reverse-spliced products resolved in non-denaturing agarose gels are templates for target-DNA-primed reverse transcription (Fig. 4b). Further, genetic experiments showed that some aI1 homing events have patterns of flanking marker co-conversion consistent with mobility mechanisms involving reverse transcription of the inserted intron RNA (R. Eskes *et al.*, in preparation).

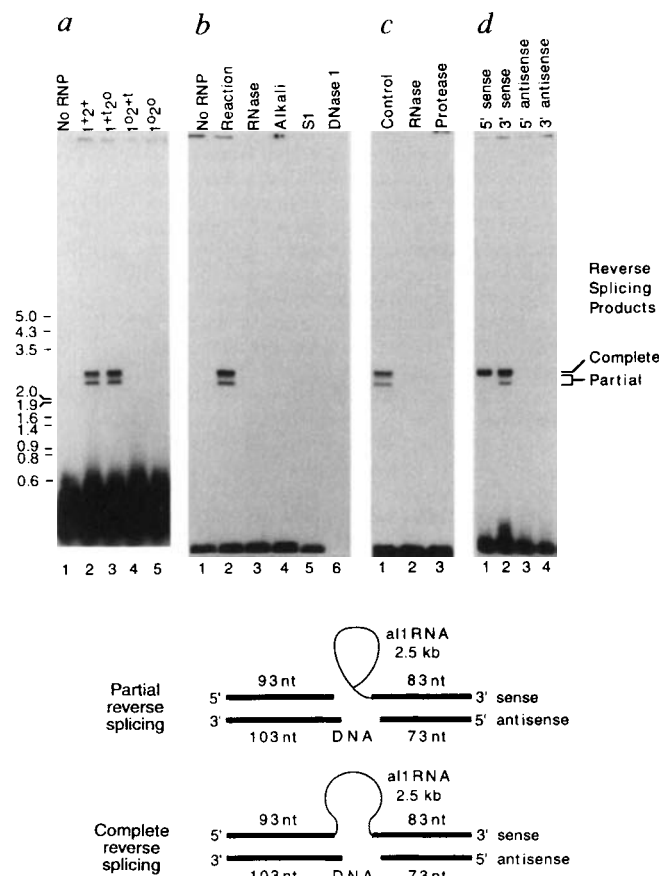


FIG. 2 Characterization of aI1 reverse splicing into DNA. a, Reverse splicing with different yeast strains. Lane 1, internally labelled DNA substrate incubated without mtRNP particles. Lanes 2–5, DNA substrate incubated with mtRNP particles from strains 161 ($1^{\circ}2^{+}$), $1^{\circ}2^{\circ}$, $1^{\circ}2^{+}$, $1^{\circ}2^{\circ}$, $1^{\circ}2^{+}$, $1^{\circ}2^{\circ}$. Numbers to the left are size markers (in kb). b, Nuclease and alkali sensitivity of reverse-spliced products. Lanes: 1, internally labelled DNA substrate incubated without mtRNP particles; 2, control reaction with $1^{\circ}2^{\circ}$ mtRNP particles; 3–6, reverse-spliced products treated with nucleases or alkali as described⁸. c, Preincubation of $1^{\circ}2^{\circ}$ mtRNP particles with RNase or protease before incubation with internally labelled DNA substrate. d, Reverse splicing reactions with $1^{\circ}2^{\circ}$ mtRNP particles and DNA substrates end-labelled at each of the four termini. The products expected for partial or complete reverse splicing of aI1 into DNA are shown.

METHODS. Reverse splicing reactions were with mtRNP particles (0.025 absorbance units at 260 nm) and ^{32}P -labelled DNA substrates (150,000 c.p.m., ~125 fmol) under the conditions used for DNA endonuclease assays (Fig. 1). Products were denatured with glyoxal and analysed in 1% agarose gels¹⁶. The internally labelled substrate was a gel-purified 176-bp DNA generated from pE1E2 by PCR using primers SK and KS (Fig. 1) in the presence of [α - ^{32}P]dGTP. The 5'-end-labelled substrates were synthesized using 5'-end-labelled SK with unlabelled E2RI (5'-ATTCTGTCAGCCAGAAAATCATT) or 5'-end-labelled E2RI with unlabelled SK. The 3'-end-labelled substrates were synthesized by filling in EcoRI- or NotI-digested pE1E2 with [α - ^{32}P]dTTP or [α - ^{32}P]dCTP, respectively, using Klenow polymerase (Life Technologies)¹⁶. After digestion with NotI or EcoRI, the 3' end-labelled 142-bp DNAs were gel-purified.

FIG. 3 al1 RNA fully reverse-splices into double-stranded DNA. **a**, Polyacrylamide gel electrophoresis of reverse-spliced products alongside *in vitro*-generated al1 lariat. Lanes: 1, *in vitro* transcript pSH2/EcoRI, containing al1 and flanking exon sequences; 2, al1 lariat RNA obtained by self-splicing of transcript pSH2/EcoRI; 3, 4, products of al1 reverse splicing with internally labelled or 5' sense-strand-labelled DNA substrates, respectively. **b**, RT-PCR strategy for analysis of intron-exon junctions in fully reverse-spliced product. **c**, Agarose gel analysis of RT-PCR products. Plus and minus signs refer to RNase treatment of the sample before RT-PCR. Lanes: M, 100-bp marker ladder (Life Technologies); 1, 2, RT-PCR to detect the E1/al1 junction; 3, 4, RT-PCR to detect the al1/E2 junction; 5–8, control RT-PCR using primers SK and al1-A1 (lanes 5, 6) or KS and al1-A2 (lanes 7, 8), with a parallel slice from a gel lane containing phenol-extracted mtRNP particles and DNA substrate; 9, 10, control RT-PCR using primers SK and al1-A1 with the more rapidly migrating partially reverse-spliced product. **d**, 5' and 3' intron-exon junctions in clones obtained via RT-PCR. Exon and intron sequences are shown in upper and lower case, respectively. The same sequence was obtained for 26 5'-junction clones and 14 3'-junction clones.

METHODS. Reverse splicing was carried out using 1⁺2⁺ mtRNP particles with internally labelled or 5' sense-strand-labelled DNA substrate, and analysed in a denaturing 3.5% polyacrylamide gel. The *in vitro* transcript containing al1 was synthesized from pSH2 linearized with EcoRI and spliced to yield al1 lariat¹⁷. Gel-purified reverse-spliced products were analysed by RT-PCR. cDNA synthesis was with 10 pmol primer KS or al1-A1 (5'-GTTAGCCCC-TACTGAGTTAGTCATA) and 200 units Superscript reverse transcriptase (Life Technologies). PCR was in 100-μl reaction medium containing 50 pmol primers SK and al1-A1 (5' junction) or al1-A2 (5'-ATACTCAATATG-GAAAGCCGTATG) and KS (3' junction) for 35 cycles (at 94 °C for 30 s; 55 °C, 30 s; and 72 °C, 30 s). RT-PCR products were cloned into pKS+ and sequenced using primers E1JY (5' junction) or al1-A2 (3' junction).

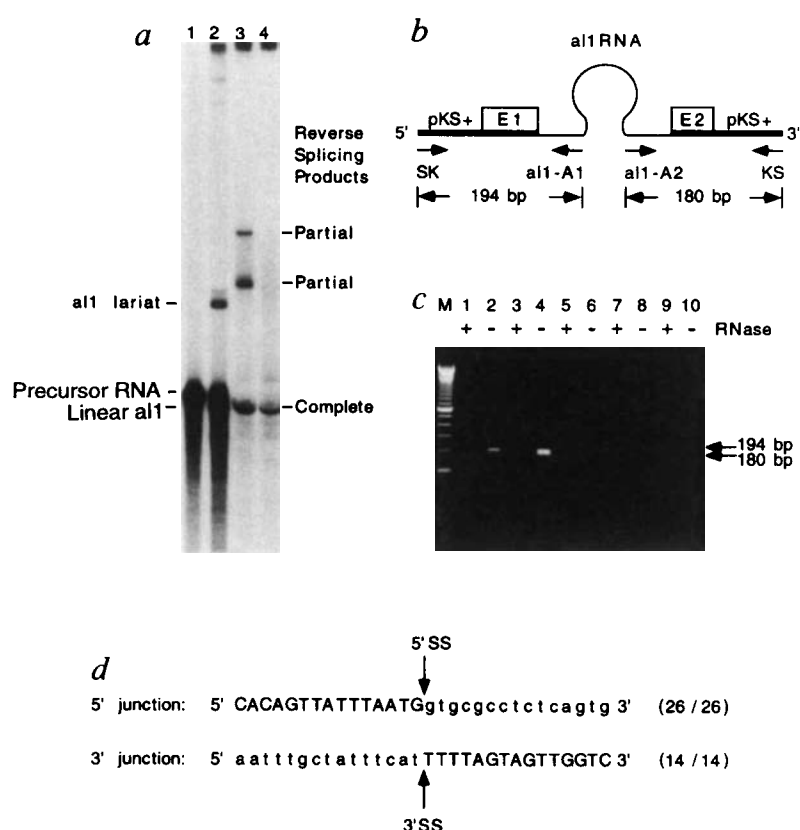
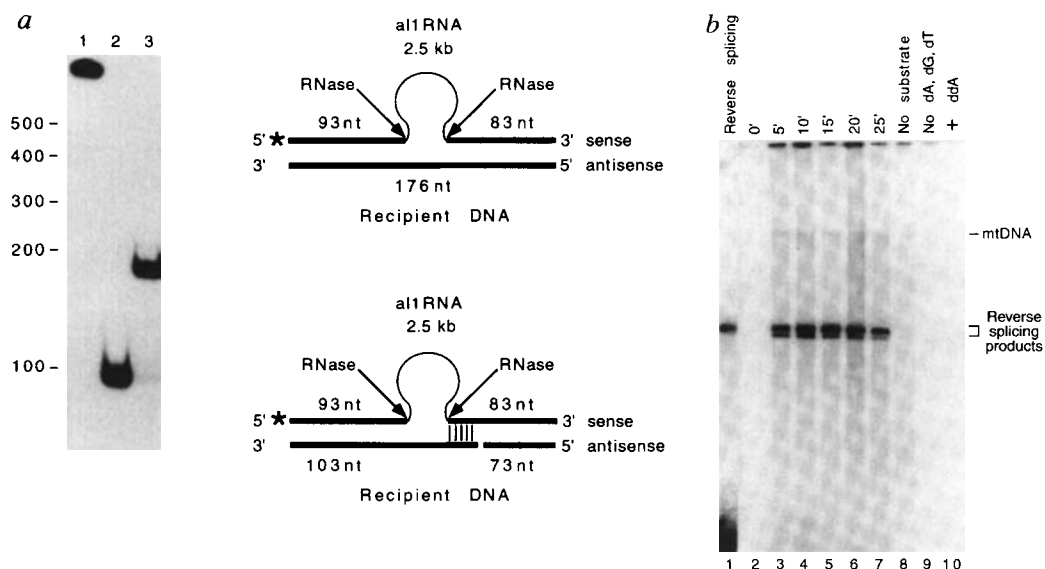


FIG. 4 DNA containing the reverse-spliced intron is cleaved on the antisense strand and is a substrate for target-DNA-primed reverse transcription. **a**, Assay of antisense-strand cleavage. Reverse splicing reactions were carried out with 1⁺2⁺ mtRNP particles and 5' sense-strand-labelled substrate (Fig. 2d). The fully reverse-spliced product was purified from a 1% agarose gel, subjected to different treatments, and then analysed in a non-denaturing 5% polyacrylamide gel to resolve double-stranded DNA products. Lane 1, 5' sense-strand-labelled reverse-spliced product (untreated); lanes 2 and 3, RNase-digested reverse-spliced product with or without incubation at 60 °C for 2 min to denature the 10-bp overlap between the halves. Numbers to the left indicate size markers (bp). The diagram shows the RNase digestion products expected for the fully reverse-spliced product cleaved or not cleaved on the antisense strand. A star indicates the location of the ³²P label. **b**, Assay of target DNA-primed reverse transcription. Wild-type 161 (1⁺2⁺) mtRNP particles were incubated with a 105-bp unlabelled DNA substrate in the presence of [³²P]dCTP and other dNTPs to support reverse transcription. Lanes: 1, reverse-spliced product synthesized using mtRNP particles and 5' sense-strand labelled DNA substrate; 2–7, time course of target DNA-primed reverse-transcription reaction; 8–10, control reverse-transcription reactions (25 min) in the absence of DNA



substrate or unlabelled dNTPs or in the presence of ddATP (100 μM). **METHODS.** Conditions for target DNA-primed reverse-transcription reactions were as described⁷. DNA substrates were synthesized from pE1E2 using primers E1JY and E2JY. MtRNP particles were incubated with unlabelled DNA substrate (600 fmol) in reaction medium containing 10 mM MgCl₂ and [³²P]dCTP plus other dNTPs. At 5 min the reaction was diluted to 2.5 mM MgCl₂ to enhance processivity of reverse transcription⁷, and at 15 min the reaction was chased with 200 μM dCTP. Products were analysed in a non-denaturing 1% agarose gel followed by autoradiography.

Although aI1 and aI2 are closely related, they are inserted at different DNA target sites, which are not recognized by the other's endonuclease⁷ (Fig. 1a, b). Comparison of the target sites (Fig. 1c) shows that the 5' exon sequences are not similar, whereas the 3' exon sequences have an 8/10 match between the intron insertion site and the antisense-strand cleavage site at position +10. The non-conserved 5' exon sequences include the IBS1 and IBS2 sequences shown to base-pair with intron sequences EBS1 and EBS2 to specify the 5' splice site in RNA forward and reverse splicing of other group II introns^{13,14}. As previous results for aI2 showed that the IBS sequences together with additional sequences are required for reverse splicing into DNA⁸, it seems likely that the different DNA-target specificities for aI1 and aI2 reflect in part a recognition of the different IBS(DNA) sequences by the intron RNA components of the endonucleases (Fig. 1c). Recent studies show that aI1 RNP particles can also reverse-splice to some extent into related, ectopic mtDNA sequences that are sites of aI1 transposition *in vivo* (J. Yang *et al.*, in preparation; compare with ref. 15).

Finally, the ability of group II introns to insert site-specifically into double-stranded DNA might provide the basis for construction of novel vectors for genetic engineering. Additional genetic information might be inserted in the domain-IV loop region or elsewhere outside the conserved RNA structure. Moreover, the intron insertion site might potentially be changed by altering the

EBS sequences to pair with new IBS sequences at a DNA target site. Thus, modified forms of the intron might be used to insert genetic information at preselected sites in DNA genomes. □

Received 1 December 1995; accepted 19 March 1996.

1. Lambowitz, A. M. & Belfort, M. A. *Rev. Biochem.* **62**, 587–622 (1993).
2. Michel, F. & Ferat, J.-L. A. *Rev. Biochem.* **64**, 435–461 (1995).
3. Meunier, B., Tian, G.-L., Macadre, C., Slonimski, P. P. & Lazowska, J. in *Structure, Function and Biogenesis of Energy Transfer Systems* (eds Quagliariello, E., Papa, S., Palmieri, F. & Saccone, C.) 169–174 (Elsevier Scientific, Amsterdam, 1990).
4. Kennell, J. C., Moran, J. V., Perlman, P. S., Butow, R. A. & Lambowitz, A. M. *Cell* **73**, 133–146 (1993).
5. Lazowska, J., Meunier, B. & Macadre, C. *EMBO J.* **13**, 4963–4972 (1994).
6. Moran, J. V. *et al. Molec. cell. Biol.* **15**, 2828–2838 (1995).
7. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. *Cell* **82**, 545–554 (1995).
8. Zimmerly, S. *et al. Cell* **83**, 529–538 (1995).
9. Moran, J. V. *et al. Nucleic Acids Res.* **22**, 2057–2064 (1994).
10. Herschlag, D. & Cech, T. R. *Nature* **344**, 405–409 (1990).
11. Robertson, D. L. & Joyce, G. F. *Nature* **344**, 467–468 (1990).
12. Mörl, M., Niemer, I. & Schmelzer, C. *Cell* **70**, 803–810 (1992).
13. Mörl, M. & Schmelzer, C. *Cell* **60**, 629–636 (1990).
14. Mörl, M. & Schmelzer, C. *Nucleic Acids Res.* **18**, 6545–6551 (1990).
15. Mueller, M. W., Allmaier, M., Eskes, R. & Schweyen, R. J. *Nature* **366**, 174–176 (1993).
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory, New York, 1989).
17. Hebbbar, S. K., Belcher, S. M. & Perlman, P. S. *Nucleic Acids Res.* **20**, 1747–1754 (1992).

ACKNOWLEDGEMENTS. We thank J. Williams for technical assistance. This work was supported by NIH grants to A.M.L. and P.S.P.

CORRESPONDENCE and requests for materials should be addressed to A.M.L. (e-mail: lambowitz.1@osu.edu).

X-ray and NMR structure of human Bcl-x_L, an inhibitor of programmed cell death

Steven W. Muchmore*, Michael Sattler†, Heng Liang†, Robert P. Meadows‡, John E. Harlan†, Ho Sup Yoon†, David Nettlesheim†, Brian S. Chang§, Craig B. Thompson§, Sui-Lam Wong¶, Shi-Chung Ng|| & Stephen W. Fesik†

* Protein Crystallography, † NMR Research, ‡ Research Computing and Information Sciences and || Aging and Degenerative Diseases Research, Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

§ Howard Hughes Medical Institute, and Departments of Medicine, Molecular Genetics and Cell Biology, The University of Chicago, Chicago, Illinois 60637, USA

The Bcl-2 family of proteins regulate programmed cell death by an unknown mechanism¹. Here we describe the crystal and solution structures of a Bcl-2 family member, Bcl-x_L (ref. 2). The structures consist of two central, primarily hydrophobic α -helices, which are surrounded by amphipathic helices. A 60-residue loop connecting helices α 1 and α 2 was found to be flexible and non-essential for anti-apoptotic activity. The three functionally important Bcl-2 homology regions (BH1, BH2 and BH3)^{3–5} are in close spatial proximity and form an elongated hydrophobic cleft that may represent the binding site for other Bcl-2 family members. The arrangement of the α -helices in Bcl-x_L is reminiscent of the membrane translocation domain of bacterial toxins, in particular diphtheria toxin and the colicins⁶. The structural similarity may provide a clue to the mechanism of action of the Bcl-2 family of proteins.

The structure of Bcl-x_L was determined by both X-ray crystal-

lography and nuclear magnetic resonance (NMR) spectroscopy (Table 1). The structural studies were conducted on a biologically active form of human Bcl-x_L lacking the putative transmembrane region at the carboxy terminus (Fig. 1 legend). Recombinant Bcl-x_L is monomeric, as determined by analytical ultracentrifugation, the lack of contact between molecules related by crystallographic two-fold axes in the unit cell, and the line widths observed in the NMR spectra (data not shown). Using initial electron-density maps generated by multiple isomorphous replacement (MIR) and density-modification techniques, approximately 60% of the residues were modelled. Fitting the missing residues to the electron-density maps was guided by the secondary structure determined from NMR data and nuclear Overhauser enhancements (NOEs) observed between α -helices. Once a model was obtained from the combined structural information, the crystal and solution structures were determined separately using only the X-ray and NMR data, respectively. Although structures of Bcl-x_L could have been obtained independently by using either technique, the structure determination proceeded more rapidly by combining information from the two complementary methods. Both the X-ray and NMR structures are well defined (Table 1; Fig. 1a, b) and agree well with each other (Fig. 1b). The root-mean-square deviation (r.m.s.d.) for the α -carbon atoms (residues 6–18 and 85–195) between the crystal structure and averaged, energy-minimized NMR structure is 1.6 Å. The differences observed between the two structures for a region following the first helix (residues 19–27) and the loop between helices α 5 and α 6 (residues 156–160) appear to be due to crystal packing. Residues 101–103 differ in the X-ray and NMR structures, but these residues are part of a flexible hydrophilic loop, and are poorly defined by both techniques.

The structure of Bcl-x_L consists of two central α -helices (α 5 and α 6), which are about 30 Å in length, flanked on one side by α 3 and α 4 and on the other by α 1, α 2 and α 7 (Fig. 1c). The two central helices contain predominantly hydrophobic residues, and are arranged in an antiparallel fashion, with a crossing angle of 23°. Helix α 6 contains a kink at Pro 180 that causes a change in the direction of this helix. The C-terminal helix (α 7) is connected to α 6 by an irregular turn composed of two glycines (Gly 186 and Gly 187), which are largely conserved. In Bcl-x_L, Gly 187 adopts

¶ Present address: Department of Biological Sciences, Division of Cellular, Molecular and Microbial Biology, University of Calgary, Calgary, Alberta T2N 1N4, Canada.

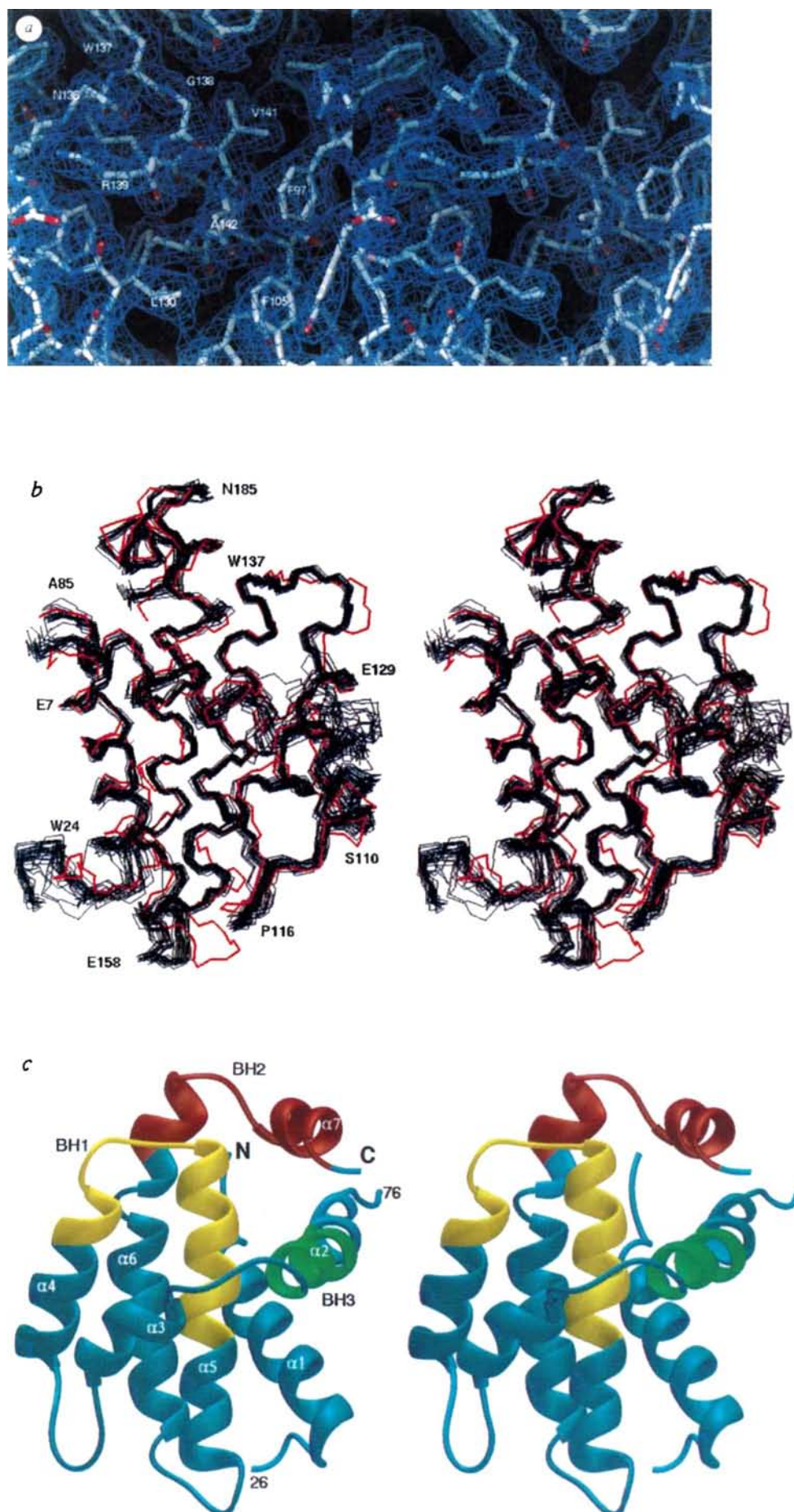


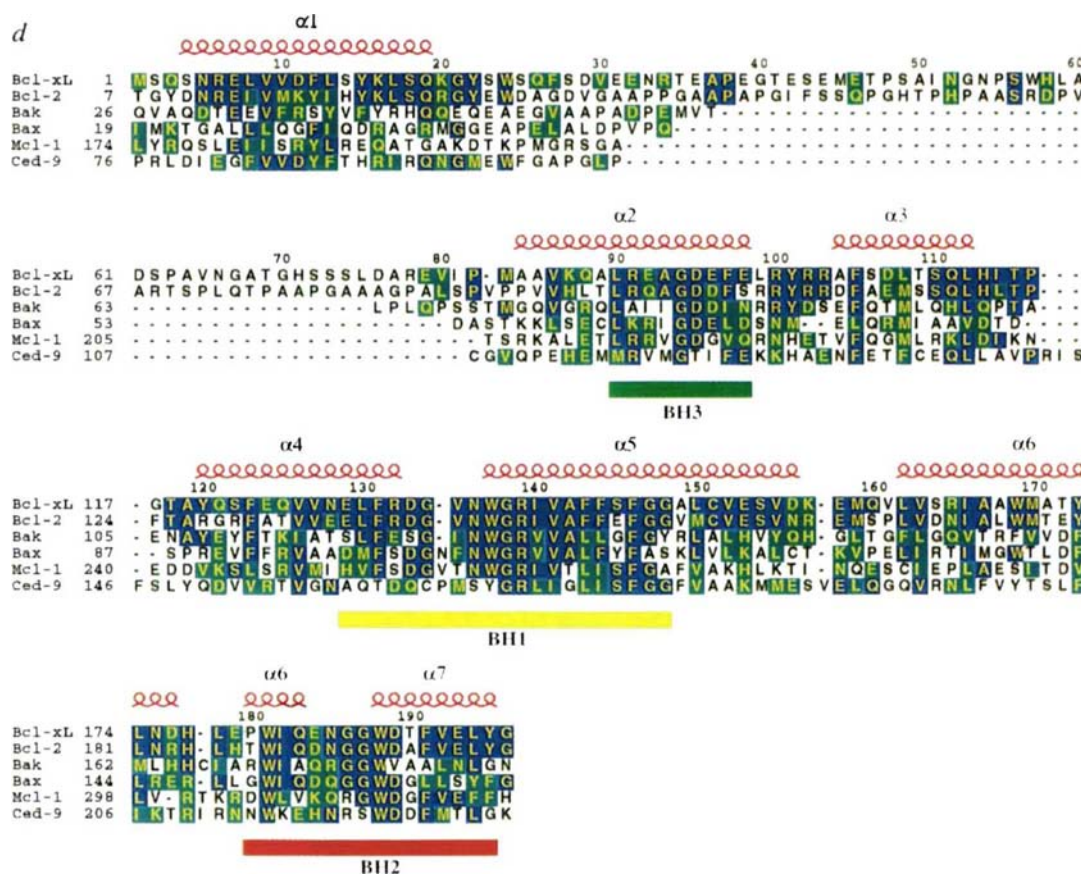
FIG. 1a, Electron density from the final $2.2\text{ \AA}$ $2F_o - F_c$ map. Blue map contours are 1 r.m.s.d. from the mean of the map. The Bcl-x_L model is shown in the stick representation, in white (C atoms), blue (N atoms) and red (O atoms). The region shown is the hydrophobic cleft near the BH1 domain. b, Stereo view of the backbone (N, C α , C') of 20 superimposed NMR-derived structures of Bcl-x_L (residues 3–26, 85–196) (black) on an N-C α -C' trace of the same residues in the X-ray structure (red). c, Ribbons¹⁵ depiction of the averaged, minimized NMR structure of Bcl-x_L. BH1, BH2 and BH3 regions are yellow, red and green, respectively. d, Sequence alignment of Bcl-2 family members. The sequence of human Bcl-x_L is shown along with those of human Bcl-2, Bak, Bax, Mcl-1 and *Caenorhabditis elegans* CED-9. Dark blue and green boxes designate residues that are identical or similar (as measured by hydrophobicity) to Bcl-x_L, respectively. The helices in the X-ray and NMR structures are indicated. The numbers above the sequences correspond to Bcl-x_L.

METHODS. Protein expression and purification: Human Bcl-x_L² (residues 1–209) lacking the C-terminal hydrophobic region was expressed in *Escherichia coli* using the pET29b vector (Novagen). The recombinant protein contained four extra amino acids (MSMA-) at the N terminus, owing to cloning artefacts, and a polyhistidine tag (-LEHHHHHH) at the C terminus to facilitate purification. To confirm that the Bcl-x_L protein analysed by both X-ray crystallography and NMR spectroscopy had biological activity, the coding region of the construct was introduced into an IL-3-dependent cell line, and transfectants survival was compared to cells transfected with wild-type Bcl-x_L or a control plasmid. For a period of 24 h after IL-3 withdrawal, the viability of control cells was $28.4 \pm 0.7\%$; transfected cells containing the construct displayed a viability of $85.0 \pm 0.4\%$, (comparable to previous¹⁶ C-terminal truncations of Bcl-x_L); and cells transfected with wild-type Bcl-x_L had a viability of $97.1 \pm 0.1\%$. These data suggest that the protein retains the ability of Bcl-x_L to function as an antiapoptotic gene. For crystallization, cells of *E. coli* strain BL21(DE3) transformed with the expression vector were grown at 37°C on LB medium. For NMR experiments, uniformly ^{15}N - or ^{15}N -, ^{13}C -labelled protein was prepared by growing the *E. coli* strain HMS174(DE3) overexpressing Bcl-x_L on M9 medium containing $^{15}\text{NH}_4\text{Cl}$ with or without $[\text{U-}^{13}\text{C}]$ -glucose. A uniformly ^{15}N -, ^{13}C -labelled and fractionally deuterated protein sample was also prepared by growing cells with $75\% \text{ } ^2\text{H}_2\text{O}$. Recombinant Bcl-x_L was purified by affinity chromatography on a nickel-IDA column (Invitrogen). Further purification was achieved by reverse-phase chromatography of the denatured protein on a C4 or C8 column, followed by refolding through the gradual removal

ϕ/ψ angles not normally observed for other residue types. This could explain its conservation (Fig. 1d) and the loss of binding to Bax on mutation of Gly 187 in Bcl-x_L (ref. 7). The N-terminal helix forms extensive hydrophobic interactions with $\alpha 2$, $\alpha 5$ and $\alpha 6$, and seems to be important for stabilizing the structure of the protein. This is consistent with the observation that deletion mutants in this region of Bcl-2 fail to protect against cell death induced by growth-factor withdrawal or to suppress Bax-mediated cytotoxicity^{8,9}.

No electron density was observed for residues Ser 28 to Val 80.

In addition, no medium- or long-range NOEs were detected for the residues in this region, and the chemical shifts and vicinal coupling constants ($^3J_{\text{NH},\text{H}_2}$) are consistent with those in a random coil. To obtain a direct measure of protein flexibility, two-dimensional ^{15}N - ^1H heteronuclear NOE experiments were conducted on uniformly ^{15}N -labelled Bcl-x_L (Fig. 2a). The negative NOE values observed for residues 35–78 (Fig. 2a, b) indicate that this region is highly mobile in solution relative to the overall tumbling rate of the molecule. This flexible loop corresponds to a region of low sequence homology and variable length among Bcl-2



of the denaturant by dialysis. X-ray crystallography: Crystals of Bcl-x_L were obtained by screening limited factorial crystallization solutions (Hampton Research). Crystals suitable for diffraction analysis were obtained by hanging-drop vapour diffusion. Protein was concentrated to 25 mg ml⁻¹ in 50 mM Tris, pH 7.4, and equilibrated against 28–31% (w/v) PEG 4,000, 0.2 M ammonium acetate, 0.1 M sodium citrate, pH 7.5. The crystals are tetragonal, space group $P4_12_12_1$, $a = b = 63.5$ Å, $c = 110.5$ Å, $\alpha = \beta = \gamma = 90.0^\circ$, and contain one copy of the Bcl-x_L protein per asymmetric unit. Diffraction data were collected at 110 K using Cu-K α radiation, on an R-Axis IIc area detector (Molecular Structure Corporation). Intensity data were evaluated using the HKL¹⁷ suite of programs. Heavy-atom difference Patterson maps were interpreted by inspection to derive the heavy-atom substituent positions. These positions were refined using MLPHARE¹⁸, and phase refinement by real-space density modification was computed using DM¹⁸. Initial model-building studies used 2.7 Å resolution electron-density maps, interpreted with the graphics program O¹⁹. Partial models were used to calculate phases that were then combined with the MIR phase set using SIGMAA²⁰ to produce improved electron-density maps. After models that agreed well with both the NMR and X-ray data were available, a final phase refinement experiment was calculated using a mask derived from the consensus structure. The final phase refinement extended to 2.2 Å resolution, and maps derived from these experiments were used to build a model used in computational refinement using X-PLOR²¹. NMR spectroscopy: NMR samples contained 1–3 mM Bcl-x_L in 20 mM sodium phosphate (pH 7.3), 5 mM perdeuterated DTT, 1 mM

perdeuterated EDTA, H₂O/²H₂O (9/1) or ²H₂O. NMR spectra were acquired at 30 °C on a Bruker DMX500 or AMX600 NMR spectrometer. The ¹H, ¹³C and ¹⁵N resonances of the backbone were assigned from a series of deuterium-decoupled triple-resonance experiments (HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB)²² recorded using the uniformly ¹⁵N-, ¹³C-labelled and fractionally deuterated protein. The side-chain signals were assigned from three-dimensional (3D) HCCH total correlation spectroscopy (TOCSY)²³ and HBHA(CO)NH²³ experiments on [U-¹⁵N-, ¹³C]-labelled samples and from HC(CO)NH-TOCSY²⁴ experiments on the [U-¹⁵N-, ¹³C]-labelled and fractionally deuterated sample. Stereospecific assignment of methyl groups of Val and Leu residues were obtained using a fractionally ¹³C-labelled sample²⁵. Distance restraints were obtained from ¹⁵N- or ¹³C-resolved 3D NOESY experiments²³ and a ¹³C-separated, ¹³C-filtered 3D NOESY acquired on a ¹²C-Phe, [U-¹⁵N-, ¹³C] sample²⁶. The ϕ angle restraints were based on the $^3J_{\text{HNH}_2}$ coupling constants measured in a HNHA experiment²⁷. Slow-exchanging amide protons were identified from a series of ¹H/¹⁵N HSQC spectra recorded after ²H₂O was added to a lyophilized ¹⁵N-labelled Bcl-x_L sample. Structures of Bcl-x_L were calculated with a distance geometry/simulated annealing protocol²⁸ using X-PLOR²¹. Structure calculations used 2,086 proton/proton distance restraints, 108 hydrogen-bond distance restraints derived for 54 slow-exchanging amide protons, and 60 ϕ angle restraints. NOE-derived restraints were categorized as strong (1.8–3 Å), medium (1.8–4 Å), or weak (1.8–5 Å) based on the observed NOE intensities.

TABLE 1 Summary of Bcl-x_L structure determination

(a) X-ray diffraction data statistics						
Crystal*	Resolution (Å)	Unique reflections†	R_{merge} (%)‡	R_{iso} (%)§	Number of sites	Phasing power
Native	2.2	11,139 (94, 80)	5.5, 22.2			
EMTS	2.6	7,233 (97, 95)	5.6, 16.1	16.6, 19.7	1	1.42, 0.93
PIP	3.5	7,004 (99, 98)	3.3, 7.4	16.2, 16.7	3	0.92, 0.61
K ₂ PtCl ₄	2.7	6,638 (92, 94)	5.6, 19.7	16.1, 12.3	2	1.51, 1.33
TMLA	2.7	7,250 (97, 94)	5.2, 18.4	26.5, 20.9	3	1.38, 1.15
(b) Structural statistics of refined X-ray model						
Total reflections (7–2.2 Å)	Working set:	9,995	Test set:	1,144		
Number of non-H atoms	Protein:	1,145	Solvent:	154		
R -factors	Standard:	19.2	Free:	25.9		
Average B -factors (Å ²)	Main chain:	16.8	Side chain:	22.3		
Stereochemical deviations	Bond lengths:	0.014	Bond angles:	1.92°		
(c) Structural statistics of NMR models						
R.m.s.d. from experimental distance restraints (Å)¶		$\langle \text{SA} \rangle$	$\langle \overline{\text{SA}} \rangle_r$			
Intra-residue (607)		0.007 ± 0.002	0.008			
Sequential (598)		0.012 ± 0.004	0.017			
Medium-range (379)		0.018 ± 0.003	0.014			
Long-range (502)		0.010 ± 0.002	0.009			
Hydrogen bonds (108)		0.030 ± 0.004	0.024			
R.m.s.d. from experimental torsional angle restraints (deg)#						
Φ angles (60)		0.23 ± 0.09	0.11			
X-PLOR potential energies (kcal mol ⁻¹)						
	$\langle \text{SA} \rangle$	$\langle \overline{\text{SA}} \rangle_r$		$\langle \text{SA} \rangle$	$\langle \overline{\text{SA}} \rangle_r$	
E_{tot}	219 ± 12	271	E_{bond}	17 ± 1	22	
E_{ang}	133 ± 5	164	E_{imp}	20 ± 1	20	
E_{repe}	27 ± 4	45	E_{noe}	20 ± 3	19	
E_{cdih}	0.2 ± 0.1	0.05	$E_{\text{LJ}}^{*\dagger}$	−1,023 ± 16	−952	
Cartesian coordinate r.m.s.d. (Å)						
$\langle \text{SA} \rangle$ versus $\langle \overline{\text{SA}} \rangle$		N, Cα and C'		all heavy atoms		
Residues 6–18, 85–195		0.61 ± 0.12		1.26 ± 0.15		
Residues 6–18, 85–100, 104–195		0.54 ± 0.10		1.03 ± 0.12		

$\langle SA \rangle$ is the ensemble of 20 simulated-annealing structures; $\langle \overline{SA} \rangle$ is the mean atomic structure obtained by averaging the coordinates of the individual $\langle SA \rangle$ following a least-squares superposition of the backbone heavy atoms (N, Cα, C') for residues 6–18 and 85–195; $\langle \overline{SA} \rangle_r$ is the energy-minimized, averaged structure. The X-PLOR F_{repe} function was used to simulate van der Waals interactions with a force constant of 4.0 kcal mol⁻¹ Å⁻⁴ with atomic radii set to 0.8 times their CHARMM values.

* Bcl-x_L crystals were derivatized by soaking crystals in stabilization solution (22% PEG 4,000, 200 mM ammonium chloride, 50 mM Tris, pH 7.5) in which the following compounds had been dissolved: EMTS, 3 mM sodium ethylmercurithiosalicylate, 24 h; PIP, 4 mM di-μ-iodobis(ethylenediamine) diplatium (II) nitrate, 48 h; K₂PtCl₄, 5 mM potassium tetrachloroplatinate (II), 24 h; TMLA, 10 mM trimethyl lead acetate, 7 days.

† The total number of merged reflections; the two numbers following the total are the overall percentage observed, and the percentage observed in the last reflection shell (~0.1 Å width).

‡ $R_{\text{merge}} = \sum_i \sum_h |I_{hi} - \langle I_{hi} \rangle| / \sum_h \sum_i |I_{hi}|$, where h specifies unique reflection indices, and i indicates symmetry equivalent observations of h and $\langle I_{hi} \rangle$ is the mean value. The first number is the value for all data, and the second refers to the highest resolution shell.

§ $R_{\text{iso}} = \sum_p \|F_{ph}\| - |F_p| / \sum_h |F_p|$, where $|F_{ph}|$ and $|F_p|$ are the measured structure factor amplitudes of the derivative and native structures. The two percentages refer to the overall and last resolution shell, respectively.

|| Phasing power is the ratio of the calculated heavy-atom structure factor amplitude to the closure error in the phasing model, and the two numbers are the overall value, and the value in the last resolution shell, respectively.

¶ NOE-derived distance restraints were used with a square-well potential ($F_{\text{noe}} = 50 \text{ kcal mol}^{-1} \text{ Å}^{-2}$). In addition, hydrogen bonds were included and given bounds of 1.8–2.4 Å (H–O) and 2.8–3.2 Å (N–O). No distance restraint was violated by more than 0.4 Å in any of the final structures. Medium-range NOEs correspond to those observed between protons separated by more than one and less than five residues in the primary sequence, and long-range NOEs are between protons separated by five or more residues.

Torsional Φ angle restraints were applied with values of $-60 \pm 40^\circ$ for $^3J_{\text{HNH}_2}$ coupling constants smaller than 5.8 Hz. These were applied only in α-helical regions. Force constants of 200 kcal mol⁻¹ rad⁻² were applied for all torsional restraints. No torsional angle restraint was violated by more than 3° in any of the final structures.

* The Lennard-Jones potential was not used during any refinement stage.

family members (Fig. 1d). To determine whether this region is required for the anti-apoptotic properties of Bcl-x_L, a mutant in which amino acids 26–83 were deleted and replaced by four alanines (HA–Bcl-x_L Δ26–83) was cloned into an expression vector and stable transfectants were analysed for their ability to survive interleukin-3 (IL)-3 withdrawal in comparison to cells

transfected with the wild-type gene (HA–Bcl-x_L). The survival of cells transfected with the deletion mutant was at least as good as the survival of cells stably transfected with full-length Bcl-x_L (Fig. 2c). Deletions of smaller numbers of amino acids in Bcl-x_L (Δ26–63 and Δ46–83) also retained anti-apoptotic properties (data not shown). These data suggest that Bcl-x_L lacking the

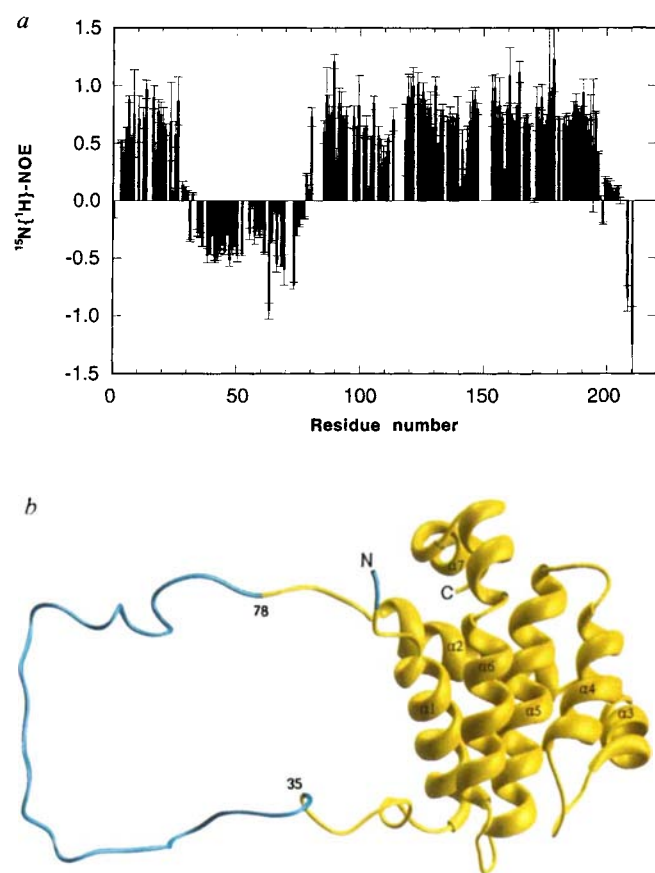
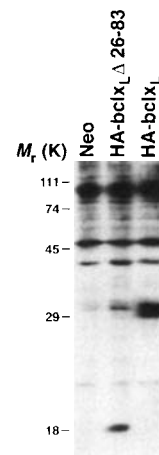
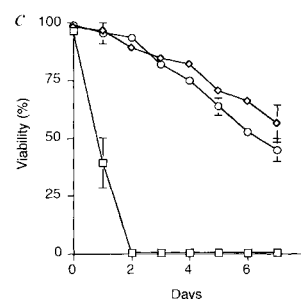


FIG. 2a, Steady-state heteronuclear $^{15}\text{N}\{^1\text{H}\}$ -NOE for the backbone amides of Bcl- x_L . b, Ribbon depiction of the energy-minimized average NMR structure. Residues exhibiting negative $^{15}\text{N}\{^1\text{H}\}$ -NOE values are blue. This schematic drawing illustrates the size of the unstructured loop relative to the rest of the protein. The conformation of the loop shown is one of many that satisfy the NMR-derived restraints. c, Functional analysis of Bcl- x_L $\Delta 26-83$ mutants. Top right, Haemagglutinin (HA)-tagged Bcl- x_L $\Delta 26-83$ (diamonds) mutant protects at least as well as the HA-Bcl- x_L (circles) against IL-3 withdrawal-induced cell death. Control transfected cells (squares) die rapidly following IL-3 withdrawal. Bottom right, western blot analysis indicating the presence of Bcl- x_L proteins in transfected cells. Arrows indicate the expected sizes of the HA-Bcl- x_L $\Delta 26-83$ (lower) or the HA-Bcl- x_L wild-type protein (upper).

METHODS. Heteronuclear $^{15}\text{N}\{^1\text{H}\}$ -NOE was measured using a pulse sequence that incorporates water flip-back, gradient coherence selection, and sensitivity-enhancement schemes²⁹. Two data sets, each consisting of spectra with proton saturation (NOE spectrum) and without proton saturation (NONOE spectrum), were collected. The $^{15}\text{N}\{^1\text{H}\}$ -NOE was calculated

large flexible loop still retains the anti-apoptotic function when overexpressed.

Bcl-2 proteins share homology domains BH1 and BH2 (Fig. 1d), and mutations within these regions in Bcl-2 or Bcl- x_L abrogate the anti-apoptotic activity and block the heterodimerization with other members of the Bcl-2 family (for example, Bax or Bak) that promote programmed cell death^{3,10,11}. A third region of homology among Bcl-2 proteins (BH3) has been recognized and found to be essential for the activity of the death-promoting proteins^{4,5}. These functionally important regions (BH1, BH2 and BH3) are in close spatial proximity (Fig. 1c), and form an elongated hydrophobic cleft in Bcl- x_L (Fig. 3a, b). This cleft may be the site of interaction with death-promoting proteins such as Bax and Bak. Indeed, mutations of a highly conserved glycine (Gly 138), which forms part of this groove (Fig. 3b, red), inhibit the death-repressor activity of Bcl-2 and Bcl- x_L and block binding



as the ratio of the heights of corresponding peaks in the NOE and NONOE spectra. Values from two data sets were averaged, error bars represent the deviation of individual values. After the NMR measurements, SDS-PAGE verified that the Bcl- x_L polypeptide was not proteolyzed. A deletion mutant of Bcl- x_L in which amino acids 26–83 have been replaced by four alanines was introduced into the pSSFV-neo expression plasmid. A HA tag was engineered on the 5' end to allow for the detection of the protein product in stable transfectants. This construct was stably introduced in FL5.12 cells by electroporation as previously described², and the survival properties of cells expressing the HA-Bcl- x_L $\Delta 26-83$ protein were compared with survival properties of cells transfected with either HA-tagged Bcl- x_L wild type or a neomycin control plasmid. At time zero, transfectants were removed from IL-3-containing media and resuspended in 2.0 ml of media lacking IL-3. Survival was then followed at daily intervals by analysing the cells for propidium iodide exclusion by flow cytometry. Each sample was analysed in triplicate; mean and s.d. are shown. The expression of HA-Bcl- x_L $\Delta 26-83$ and HA-Bcl- x_L was confirmed by western blot analysis.

to Bax³. Although this glycine was proposed to play an important structural role³, its ϕ and ψ angles (-56° and 41° , respectively) are within the range of other amino acids. Thus Gly 138 seems to be important because amino acids with bulkier side chains could potentially block access to this pocket. Further evidence to support the functional importance of this region is the high conservation of the hydrophobic residues (Fig. 1d) that line this cleft which are solvent exposed (Fig. 3a, b).

The sequence homology between Bcl- x_L and other members of the Bcl-2 family (Fig. 1d) suggests that these proteins exhibit a similar fold. Bcl- x_L resembles the membrane insertion domain found in bacterial toxins⁶, in particular the diphtheria toxin and the colicins. Like Bcl- x_L , these membrane insertion domains contain two central helices consisting of apolar residues, and are long enough to span a membrane. Helices $\alpha 5$ and $\alpha 6$ and three of the surrounding amphipathic helices ($\alpha 1$, $\alpha 3$ and $\alpha 4$) in the Bcl- x_L

structure align well (r.m.s.d., $2.0\text{ \AA}$ for 50 $\text{C}\alpha$ positions) with corresponding elements of the transmembrane domain of diphtheria toxin (Fig. 3c),

The diphtheria-toxin translocation domain is thought to dimerize and form a pH-dependent membrane pore¹². By analogy, Bcl-2 proteins may form pores in the cytoplasmic membranes where they localize. The insertion of Bcl-2 proteins like the bacterial

proteins, may be regulated by signals dependent on voltage and pH. Membrane insertion and pore formation might promote cell death by dissipating the essential electrochemical gradients necessary for cellular homeostasis, or alternatively could inhibit cell death by eliminating a chemical imbalance and restoring cellular homeostasis. Bcl-2 overexpression stabilizes calcium homeostasis between the cytoplasm and the ER¹³. Bcl-2 has also been shown to

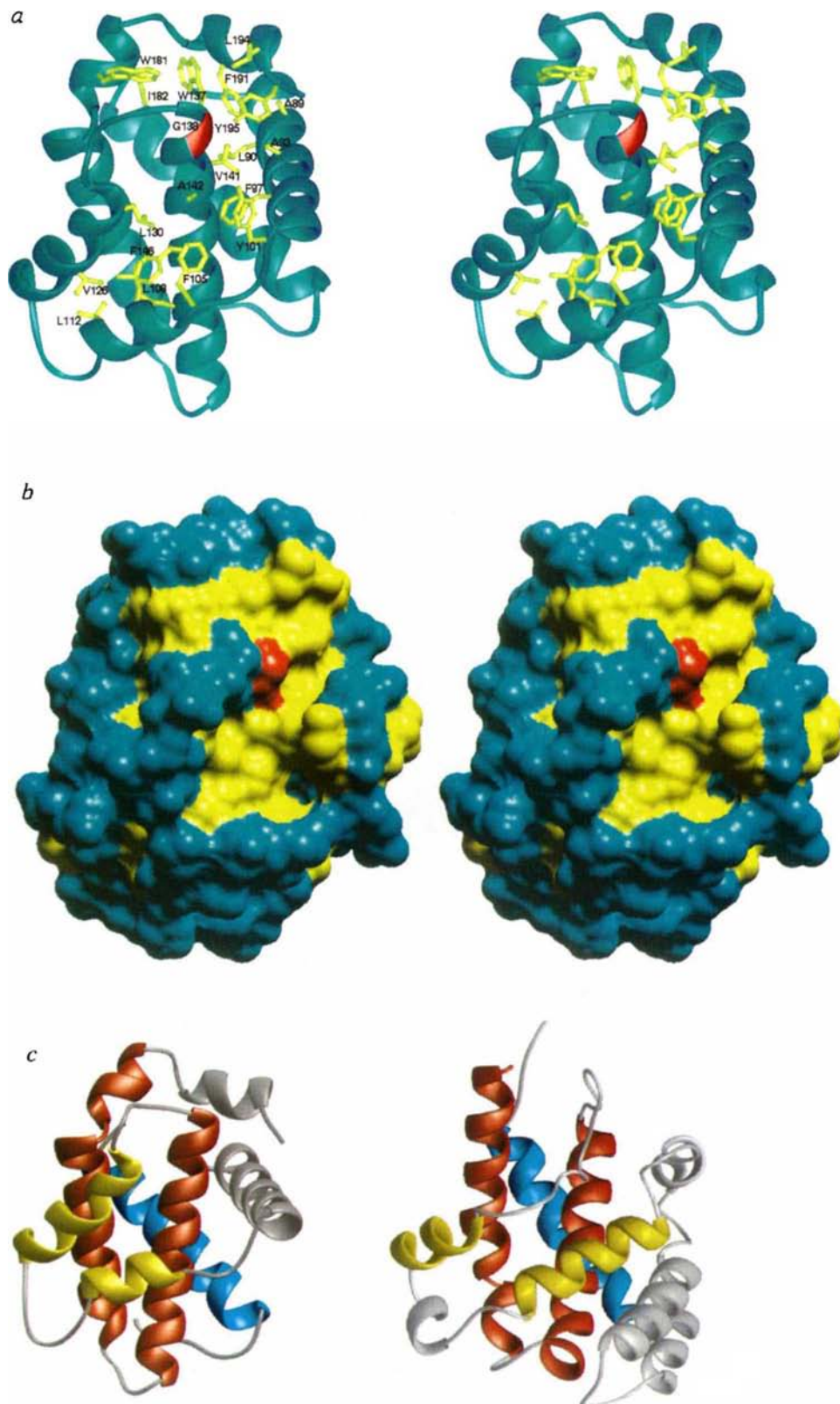


FIG. 3a, Ribbons stereo view of the X-ray structure of Bcl-x_L showing the position of hydrophobic side chains (yellow) and Gly-138 (red) in the region of the proposed binding site. b, Stereo view of a surface representation of Bcl-x_L in the same orientation as a. Ile, Leu, Val, Phe, Trp, Tyr and Ala are shown in yellow, and Gly-138 in red. c, Structural comparison of Bcl-x_L (left) and the transmembrane domain of diphtheria toxin³⁰ (right). The two central helices are shown in red, and the amphipathic α -helices common to both proteins are blue or yellow.

prevent mitochondrial permeability transition, which results in the uncoupling of electron transport¹⁴. These data suggest that Bcl-2 proteins may directly or indirectly affect the permeability of these important cytoplasmic organelles and thus regulate cellular homeostasis and apoptosis. □

Received 26 February; accepted 11 April 1996.

- Korsmeyer, S. J. *Trends Genet.* **11**, 101–105 (1995).
- Boise, L. H. *et al. Cell* **74**, 597–608 (1993).
- Yin, X.-M., Oltvai, Z. N. & Korsmeyer, S. J. *Nature* **369**, 321–323 (1994).
- Chittenden, T. *et al. EMBO J.* **14**, 5589–5596 (1995).
- Boyd, J. M. *et al. Oncogene* **11**, 1921–1928 (1995).
- Parker, M. W. & Pattus, F. *Trends biochem. Sci.* **18**, 391–395 (1993).
- Cheng, E. H.-Y., Levine, B., Boise, L. H., Thompson, C. B. & Hardwick, J. M. *Nature* **379**, 554–557 (1996).
- Borner, C. J. *Cell Biol.* **126**, 1059–1068 (1994).
- Hanada, M., Aime-Sempe, C., Sato, T. & Reed, J. C. *J. biol. Chem.* **270**, 11962–11969 (1995).
- Sedlak, T. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 7834–7838 (1995).
- Yang, E. *et al. Cell* **80**, 285–291 (1995).
- London, E. *Biochim. biophys. Acta* **1113**, 25–51 (1992).
- Lam, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 6569–6573 (1994).
- Zamzami, *et al. J. exp. Med.* **182**, 367–377 (1995).
- Carson, M. J. *J. Molec. Graph.* **5**, 103–106 (1987).

- Fang, W., Rivard, J. J., Mueller, D. L. & Behrens, T. W. *J. Immunol.* **153**, 4388–4398 (1994).
- Otinowski, Z. *Proceedings of the CCP4 Study Weekend: Isomorphous Replacement and Anomalous Scattering* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (SERC Daresbury Laboratory, Warrington, UK, 1991).
- CCP4: A Suite of Programs for Protein Crystallography (SERC Daresbury Laboratory, Warrington, UK, 1979).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr. A* **47**, 110 (1991).
- Reed, R. J. *Acta crystallogr. A* **42**, 140–149 (1986).
- Brünger, A. T. *X-PLOR 3.1* (Yale University, New Haven, CT, 1992).
- Yamazaki, T., Lee, W., Arrowsmith, S. H., Muhandiram, D. R. & Kay, L. E. *J. Am. chem. Soc.* **116**, 11655–11666 (1994).
- Clore, G. M. & Gronenborn, A. M. *Meth. Enzym.* **239**, 349–363 (1994).
- Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. *FEBS Lett.* **314**, 413–418 (1992).
- Neri, D., Szperski, T., Otting, G., Senn, H. & Wüthrich, K. *Biochemistry* **28**, 7510–7516 (1989).
- Vuister, G. W., Kim, S.-J., Wu, C. & Bax, A. *J. Am. chem. Soc.* **116**, 9206–9210 (1994).
- Kuboniwa, H., Grzesiek, S., Delaglio, F. & Bax, A. *J. Biomolec. NMR* **4**, 871–878 (1994).
- Kuszewski, J., Nilges, M. & Brünger, A. T. *J. Biomolec. NMR* **2**, 33–56 (1992).
- Farrow, N. A. *et al. Biochemistry* **33**, 5984–6003 (1994).
- Choe, S. *et al. Nature* **357**, 216–222 (1992).

ACKNOWLEDGEMENTS. We thank T. Holzman, J. Severin, K. Walter and H. Zhang for help in protein preparation; S. Snyder for analytical ultracentrifugation studies; and S.W.M. thanks C. Abad-Zapatero, C. Park and V. Giranda for discussions.

CORRESPONDENCE and requests for materials should be addressed to either S.W.M. (e-mail: muchmore@servo.pprd.abbott.com) or S.W.F. (e-mail: fesik@steves.abbott.com). Coordinates for the NMR and X-ray structures of Bcl-x_L have been deposited in the Brookhaven Protein Data Bank with accession numbers 1LXL and 1MAZ, respectively.

Structure of mitochondrial creatine kinase

Karin Fritz-Wolf, Thomas Schnyder*†, Theo Wallimann* & Wolfgang Kabsch

Max-Planck-Institut für medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, D-69028 Heidelberg, Germany
* Institute for Cell Biology, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

CREATINE kinase (CK; EC 2.7.3.2), an enzyme important for energy metabolism in cells of high and fluctuating energy requirements, catalyses the reversible transfer of a phosphoryl group from phosphocreatine to ADP^{1–3}. We have solved the structure of the octameric mitochondrial isoform, Mi₆-CK, which is located in the intermembrane compartment and along the cristae membranes. Mi₆-CK consumes ATP produced in the mitochondria for the production of phosphocreatine, which is then exported into the cytosol for fast regeneration of ATP by the cytosolic CK isoforms. The octamer has 422 point-group symmetry, and appears as a cube of side length 93 Å with a channel 20 Å wide extending along the four-fold axis. Positively charged amino acids at the four-fold faces of the octamer possibly interact with negatively charged mitochondrial membranes. Each monomer consists of a small α -helical domain and a large domain containing an eight-stranded antiparallel β -sheet flanked by seven α -helices. The conserved residues of the CK family form a compact cluster that covers the active site between the domains.

We have obtained tetragonal crystals of sarcomeric, chicken cardiac Mi₆-CK, both with and without NaATP, which contain half an octamer in the asymmetric unit. The native structure was solved by isomorphous replacement at 5 Å, and phase extension to 3 Å resolution by using the four-fold redundancy of the electron density (Table 1). The monomer (dimensions $\sim 36 \times 42 \times 69$ Å) consists of a small (residues 1–112) and a large domain (residues 113–380), with the substrate-binding site located in the cleft between the domains (Fig. 1). Sequence alignment of the CK isoenzymes^{4,5} reveals six highly conserved regions, which form a compact common core of the enzyme involved in substrate binding and catalysis (Fig. 1a, b). Outside the common core we have identified regions that are involved in isoform-specific functions of Mi₆-CK, such as the formation of octamers or binding to

the mitochondrial membrane. The overall chain fold differs from other kinases, although the sequence motif (L₂₈₂GTGLRAGV₂₉₀) matches the consensus sequence (LGXGX_(A,S)V) of the P-loop found as a characteristic feature in the protein-kinase family⁶. In CK it extends into strand β_6 , and is located too far from the phosphate moiety of ATP for a direct interaction (Fig. 2a, bottom). A remote similarity also exists for strands β_6 , β_7 and β_8 , which are connected like strands 3, 4 and 5 in the catalytic subunit of cyclic AMP-dependent protein kinase (see Fig. 3 ref. 6).

Data analysis of the ATP : Mi₆-CK co-crystals reveals different electron density for the nucleotide in each monomer. Structural details of the nucleotide and its protein environment, as found in monomer A (and similarly in C and D) of the ATP : Mi₆-CK co-crystals are described in Figs 1a and 2a, b. ATP is bound differently in monomer B (Figs 1b and 2c), such that the adenine ring comes closer to Trp 223. The different locations, as well as the solvent accessibility of the phosphate moiety and the high-temperature factors for all nucleotides and their surrounding loops, suggest that the nucleotides have not reached their final well-defined places required for the phosphoryl-transfer reaction. We speculate that in response to the binding of a magnesium ion and creatine (not present in the crystals), the nucleotide base would be closer to Trp 223 and the phosphates nearer to Asp 228, Glu 227 and Glu 226 (which are good candidates for coordinating the magnesium ion of ATP, as well as for binding creatine), and the loops 60–66 and 316–326 would move nearer the active site to exclude water during catalysis. Indeed, these loops are very flexible in the crystal, as indicated by solvent exposure of hydrophobic residues (Phe 63 and Ile 64), high-temperature factors, and different conformations assumed in the monomers resulting from crystal contacts. Moreover, Ala 323, which is part of the loop 316–326 (Fig. 1b), has previously been identified as a cleavage site for proteases, and was thought to be located in an exposed surface loop involved in the substrate-induced conformational changes essential for catalysis⁷.

Although none of the nucleotides found in the structure are at the correct places for catalysis, they must be quite close. The peptide regions Val 232 to Lys 237 and Val 275 to Arg 287 in the neighbourhood of the adenine base (Fig. 2a) have been photo-labelled with azidoadenine nucleotide analogues⁸, and many of the active-site residues implicated by various other methods can be identified in the X-ray structure.

Cys 278 (Cys 283 in cytosolic CKs), which is found near the γ -phosphate of ATP (Fig. 2a, b, bottom) was shown by affinity labelling with epoxycrystalline to be located in or near the creatine-binding site⁹. The distance of 10.3 Å between the S _{γ} atom of Cys 278 and the P _{β} atom of ATP is in qualitative agreement with

† Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

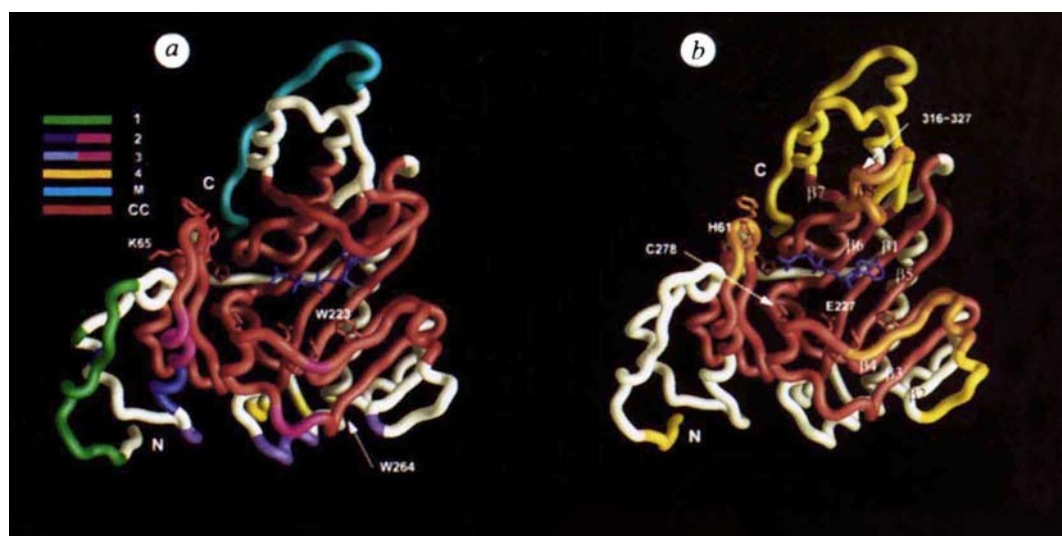


FIG. 1 Monomer structure. The course of the polypeptide chains is shown in the 'worm' representation, provided by the graphics program GRASP (A. Nicholls). The bound ATP is in blue. The common core (CC, red) of conserved amino-acid residues of CK isoenzymes includes the catalytic centre of the molecule. The conserved regions are Ile 52–His 92, Tyr 120–Ile 132, Gln 180–Lys 242, Phe 266–His 291, Arg 311–Val 325 and Asp 330–Val 352. Structural differences between the four crystallographically independent monomers in the octamer are rare, comprising residues 1–10, 60–66, 109–120, 313–327, and the position of the bound nucleotide. The structures of monomers A, C and D (a) are roughly the same but pronounced differences are found in monomer B (b). a, Assembly into octamers involves four contact regions in each monomer: contact 1 (green), Lys 5–Lys 20 and Pro 31; contact 2 (blue), Tyr 34 and Ser 47–Cys 51 (pink), Ile 52–Asn 58; contact 3 (violet), Asn 44–Gly 45, Ser 142–Arg 147 and Thr 172 (pink), Lys 191 and Arg 204–Trp 206; contact 4 (yellow), Gly 134–Ser 136, Glu 261–Trp 264. The pink regions in contacts 2 and 3 are also part of the conserved region. The region M (light blue), Lys 360–Lys 380, is involved in binding to the negatively charged head groups of lipids in the mitochondrial membranes. b, Flexible segments of the polypeptide chain of monomer B with temperature factors above $40 \text{ \AA}^2$ are yellow, or orange if they are also part of the conserved region. The locations of the flexible loops 60–65 and 314–331, and the strands β_1 to β_8 , as defined in c, are indicated. c, Schematic representation of secondary structure. First and last amino-acid residues in the helices and sheet strands are specified according to the automatic procedure DSSP³⁰. Each monomer can be partitioned into two domains: the small domain I consists of a 3_{10} -helix and five α -helices of which helix α_3 is distorted; the large domain II contains a central 8-stranded antiparallel β -sheet surrounded by seven α -helices. A second β -sheet of three very short antiparallel strands (Trp 268–Asn 269, Gly 273–Tyr 274, Gly 283) has been omitted. d, Stereo pair of a representative portion of the final $2F_o - F_c$ map covering residues Phe 189, Trp 206 and Trp 223. Density is contoured at 20% of the map maximum.

d

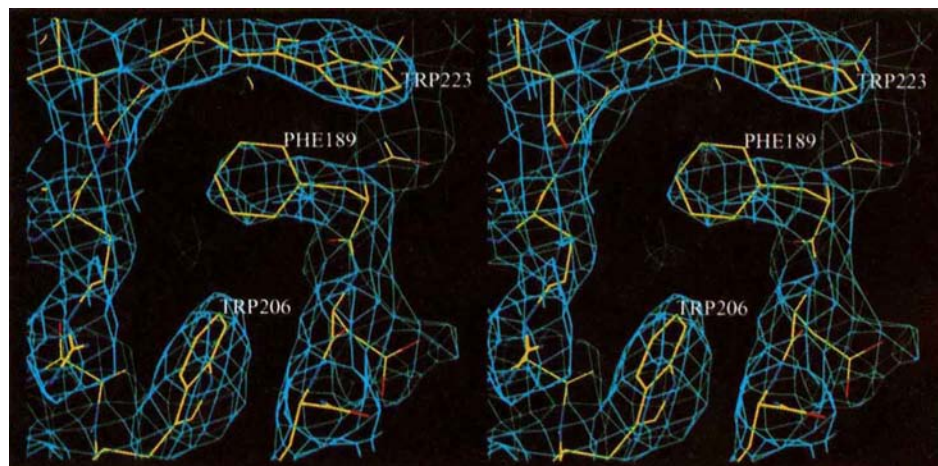


TABLE 1 Statistics of crystallographic data

Resolution bins (Å):	Overall	∞–6.0	6.0–5.0	5.0–4.0	4.0–3.5	3.5–3.0
ATP : Mi ₀ -CK (<i>a</i> = <i>b</i> = 126.0 Å, <i>c</i> = 144.8 Å)						
<i>R</i> _{sym} (%) (3 crystals)*	9.4	4.1	6.9	8.3	15.9	35.5
Mean redundancy†	3.3	3.7	3.9	3.7	3.2	2.7
Completeness (%)	96.7	93.7	99.3	98.7	98.0	95.4
of data with <i>I</i> > σ	72.3	89.1	90.8	86.7	76.0	52.0
<i>R</i> (%)‡	22.7	26.8	22.0	18.3	22.2	29.0
<i>R</i> _{free} (%)§	27.9	34.5	27.8	22.6	26.3	34.1
Mi ₀ -CK (native) (<i>a</i> = <i>b</i> = 126.0 Å, <i>c</i> = 144.7 Å)						
<i>R</i> _{sym} (%) (7 crystals)*	11.4	6.7	9.6	11.3	20.3	39.7
Mean redundancy†	6.0	9.8	7.9	7.4	6.1	3.3
Completeness (%)	99.5	99.1	100.0	100.0	100.0	99.0
of data with <i>I</i> > σ	88.0	98.5	98.4	97.7	93.6	73.3
<i>R</i> (%)‡	21.7	26.5	21.8	17.5	20.5	27.2
<i>R</i> _{free} (%)§	26.4	32.2	26.6	22.2	24.0	31.4
Mean figure of merit	0.47	0.51	0.41			
PCMB61 (<i>a</i> = <i>b</i> = 126.1 Å, <i>c</i> = 145.1 Å)						
Completeness (%)	85.1	86.8	82.6			
Phasing power	1.6	1.8	1.3			
<i>R</i> _c (%)¶	53.4	48.3	69.7			
PCMB3 (<i>a</i> = <i>b</i> = 125.8 Å, <i>c</i> = 145.0 Å)						
Completeness (%)	99.7	99.4	100.0			
Phasing power	1.6	1.9	1.2			
<i>R</i> _c (%)¶	50.4	46.4	62.9			

Sarcomeric Mi₀-CK was purified from chicken myocardium¹⁶. Enzymatic activity was verified by measuring the H⁺ consumption in the reaction^{18,24}. For crystallization, the protein solution was dialysed against buffer (in mM) (25 sodium phosphate, 1 NaN₃, 2 β-mercaptoethanol, 0.2 Na₂EDTA, pH 6.7) and diluted to 5 mg ml⁻¹. Crystals (space group *P*4₂, *a* = 126.0 Å, *c* = 144.7 Å) were grown at room temperature by the hanging-drop technique, using 17% PEG 1,000 (Merck) as precipitant in the above buffer. In the drop, the initial concentration of the protein was 2.5 mg ml⁻¹, while that of the precipitant was half the concentration in the reservoir solution. These crystals are better suited for X-ray analysis than the other forms described previously^{25,26}. Co-crystals of the new type were grown in the presence of 5 mM NaATP (Boehringer) in the drop solution. After data collection, the co-crystals were freed from unbound ATP and analysed on a Superose 12 gel-permeation chromatography column (Pharmacia) for verification that ATP was present in the crystals. X-ray diffraction images were collected at 4 °C by the rotation method and processed by the program XDS²⁷. Native and heavy-atom derivative crystal data were recorded by an electronic area detector (Siemens/Nicolet, Madison, WI; crystal-to-detector distance, 13 cm; exposure time per image, 3–4 min; rotation range per image, 0.1–0.01°) using X-rays (CuK_α-line from a rotating anode operated at 35 kV and 50 mA, GX-18 Elliott/Enraf-Nonius, Delft) focused by Franks double-mirror optics. Two heavy-atom derivative data sets (PCMB3 and PCMB61) were obtained by soaking native crystals in mother liquor containing 22% PEG 1,000 and 2 mM *p*-chloromercuribenzoate (Sigma) for 6 and 12 h at room temperature. Diffraction data from the ATP : Mi₀-CK co-crystals were recorded by an imaging plate detector (X-rayresearch, Hamburg; crystal-to-detector distance, 25 cm; rotation range per image, 1°) using synchrotron radiation (wavelength, 0.91 Å) at the EMBL-outstation (DESY). Structure determination used programs from W.K. (unpublished), and X-PLOR²⁸ and O²⁹ for refinement and model building, respectively. Analysis of the native data set, Mi₀-CK (native), by the rotation function revealed 422 point-group symmetry of the Mi₀-CK octamer. One of the two-fold axes coincides with a crystallographic two-fold axis; the four-fold symmetry axis is in the *ab* plane of the crystal at an angle of 77.3° to the *a*-axis. The unit cell contains two octamers with the centres at the special positions 0, 1/2, 0 and 1/2, 0, 1/2, respectively. Deconvolution of the difference Patterson maps of the PCMB-derivatives revealed 4 (data set PCMB61) or 8 (data set PCMB3) heavy-atom sites related by the non-crystallographic symmetry of the octamer. Phase information to 3 Å resolution was extracted from the four-fold redundancy of the density in the asymmetric unit. The complete atomic model was obtained by several rounds of model building and correction followed by refinement and map calculation. The strict non-crystallographic symmetry was slightly relaxed in the final stages of refinement as the amino-acid chain adopts different conformations in each monomer for residues 1–10, 60–66, 109–120 and 313–327. The geometry of the atomic model, which contains no water molecules, is close to standard values with an average deviation of 0.007 Å in bond distances and 1.2° in bond angles. Apart from the peptide connecting Trp 206 with Pro 207, which is probably in a *cis*-conformation, the average deviation from the ideal *trans*-conformation is 7°. Uncertainties in the structure assignments remain for loop residues 316–326, 60–66 and the first five residues at the N terminus (located in weak electron density). The ATP : Mi₀-CK structure was solved using the native model.

* $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \bar{I}_h| / \sum_h \bar{I}_h$, where *h* are unique reflection indices and *I*_{hi} are the intensities of symmetry equivalent reflections giving a mean value of *I*_h.

† Multiplicity of measurements of each unique reflection.

‡ $R = \sum |F_{\text{obs}} - F_{\text{model}}| / \sum F_{\text{obs}}$, where *F*_{obs} and *F*_{model} are observed and atomic-model structure factor amplitudes, respectively. Reflections at resolution lower than 8 Å are excluded from the summation.

§ *R*-factor calculated for 10% of the reflections excluded from the refinement.

|| Phasing power is the mean value of the heavy-atom structure factor amplitude divided by the residual lack-of-closure error.

¶ Cullis *R*-factor for centric reflections.

results of nuclear magnetic resonance (NMR) studies, from which a distance of 7–10 Å between the manganese ion of MnADP and the nitroxide moiety of the spin-label attached to Cys 278 was estimated¹⁰.

A carboxyl group, which must be ionized for the binding of both creatine and phosphocreatine, has been implicated from studies of pH rate¹¹. In the X-ray structure we find Asp 228, Glu 227 and Glu 226 (lower right in Fig. 2a) as possible candidates which, in addition, could be involved in coordinating the magnesium ion of ATP. The first two residues are also in the vicinity of Cys 278, which is thought to mark the creatine-binding site⁹, as mentioned above. Asp 335, which has been radioactively labelled by CIRATP, an ATP analogue with the reactive label covalently bound to the γ-phosphate group¹², forms a salt bridge with Arg 311, and its side chain is pointing away from Cys 278 and the phosphate moiety of

ATP (lower left in Fig. 2a), which makes it less likely that this residue is involved in creatine binding.

The presence of arginine and lysine residues in the catalytic site has been inferred from chemical modification and NMR studies (reviewed in refs 1,2). The X-ray structure reveals five arginines and one lysine in the catalytic region of the molecule. The side chain of Lys 314 is pointing away from the nucleotide, which suggests that it has a more indirect role during catalysis.

NMR studies on the role of histidines at the active site have revealed four such residues at distances 12, 12, 14 and > 18 Å from the Cr³⁺ in the paramagnetic β,γ-bidentate Cr³⁺ ATP complex bound to the enzyme¹⁵. One of these residues, with a pK near 7, was postulated to act as an acid–base catalyst¹¹. We find His 92, His 291, His 229, His 61 and His 186 at distances of 12, 12,

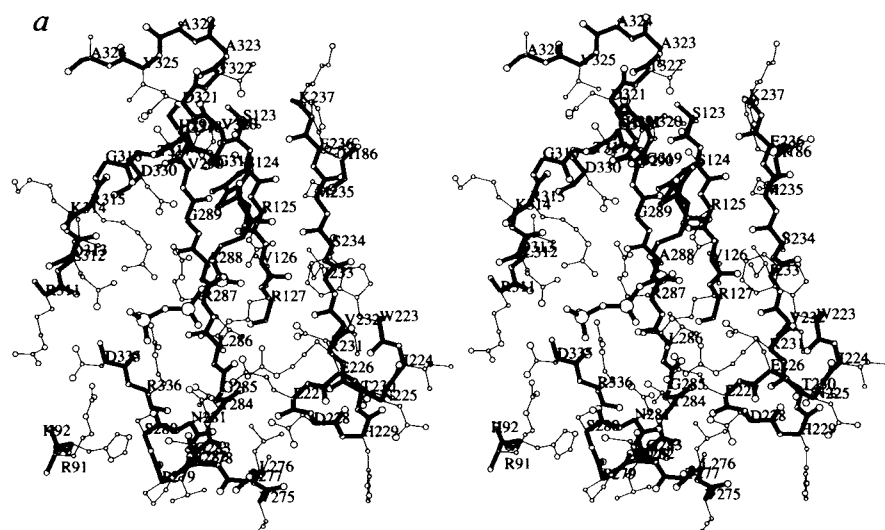
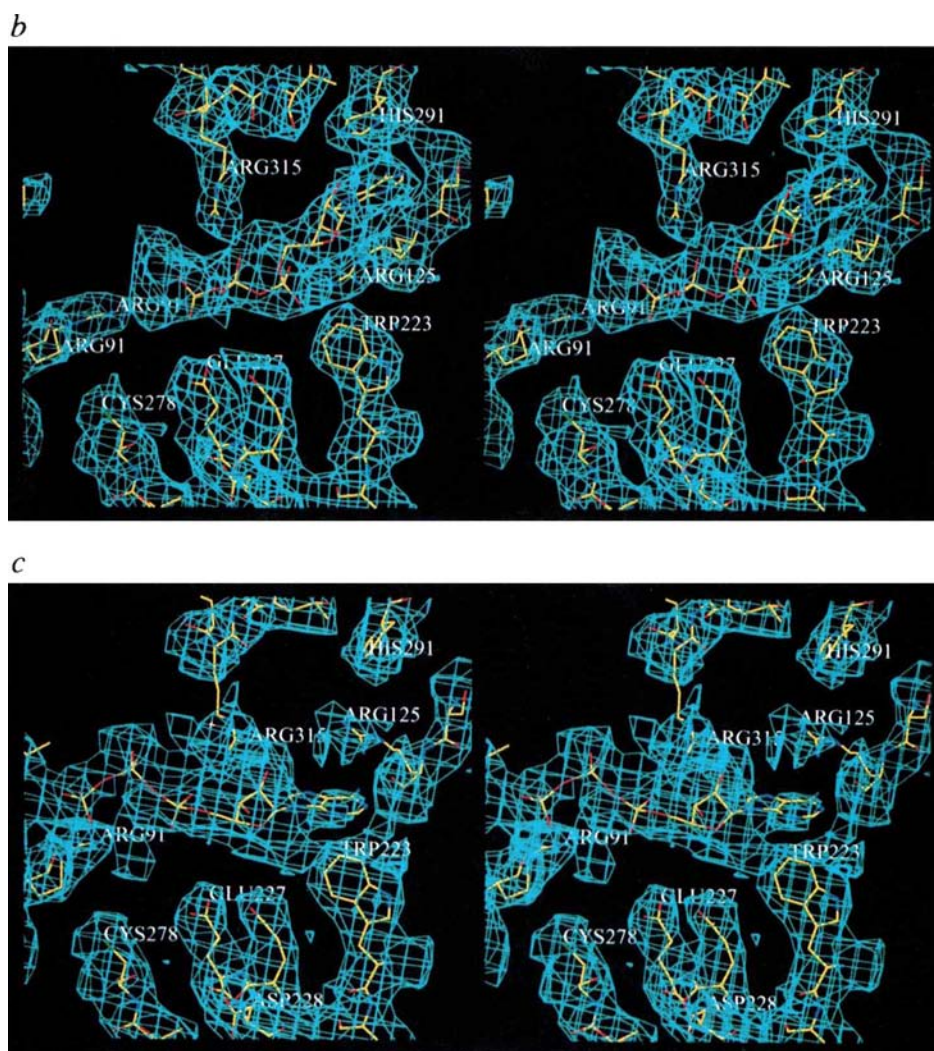


FIG. 2 Stereo pairs of the active site of Mj-CK. *a*, Atomic details of the nucleotide-binding region in monomer A (similar to monomers C and D). The adenine base fits into a pocket formed by residues Met 235, His 186, Ser 123, His 291, Gly 289, Arg 125 and the pair Arg 287, Asp 330, which forms a salt bridge. There are no specific interactions between these amino acids and the adenine base which is found in the anti orientation relative to the ribose. The dihedral angle χ of the *N*-glycosidic bond is $\sim -134^\circ$. The phosphate groups of ATP interact with five arginine residues at positions 287, 125, 127, 315 and 336, with two (125 and 127) belonging to strand β_1 , which connects the domains. An additional residue, Arg 91, is 5.4 Å from the γ -phosphate. *b*, Omit map of the nucleotide and some neighbouring residues in monomer A. Electron density was obtained using computed phases from the atomic model refined without ATP. Density is contoured at 10% of the map maximum. *c*, Omit map of the nucleotide in monomer B (same view and contour level as *b*). Consistent with the higher temperature factor, the density is less well defined than for the nucleotides in the other monomers.



14, 17 and 18 Å from the O7 atom of ATP, respectively (Fig. 2a). His 92 is closest to the γ -phosphate and to Cys 278. However, we cannot exclude that His 61, which resides on a flexible loop thought to be involved in the structural reorganization during catalysis, could also approach the active site. Thus we can only speculate that His 92 or possibly His 61 could function as the general acid-base catalyst.

The presence of a tryptophan residue in the vicinity of the adenine part of the substrate (< 5 Å) has been deduced from NMR and fluorescence-spectroscopic observations¹⁴. It was later shown by site-directed mutagenesis that replacement of Trp 223 results in a complete inactivation of the enzyme¹⁵. In the X-ray structure, Trp 223 is at a larger distance, about 10 Å, from the adenine (middle right in Fig. 2a, b) but, as mentioned above, we

expect that the nucleotide would approach Trp223 in the presence of Mg^{2+} .

Mi_b -CK octamers are quite stable, the dissociation into dimers upon dilution taking hours to weeks^{16,17}. This process is greatly accelerated by the addition of a transition-state analogue-complex mixture (TSAC = creatine, MgADP and nitrate)¹⁸, which leads to dissociation within minutes¹⁷. The dimers of all CK isoforms are very stable, and disintegrate into monomers only under chaotropic conditions. The crystal structure explains the observed stability of the octamer and its assembly from dimers by numerous interactions of its monomers (Fig. 3). Most of the contacts involve isotype-specific regions in the amino-acid sequence outside the conserved CK framework⁵. Exceptions are Ile 52 to Asn 58 in contact region 2, and residues Lys 191 and Arg 204 to Trp 206 in contact region 3, which are also conserved residues (Fig. 1a). This is consistent with the finding that these residues are involved in the formation of CK dimers, which is a common property of both the mitochondrial and cytosolic isoforms of CK. Form and size of the dimers and their orientation in the octamer agree well with results of electron microscopy¹⁹. Moreover, the contribution of Trp 206 in contact region 3 to the stability of the dimers has been shown by site-directed mutagenesis and fluorescence spectroscopy¹⁵.

Trp 264 has been suggested to be part of a hydrophobic inter-

dimer-interaction patch important for octamer stability, because replacement of this residue by cysteine drastically increases the TSAC-induced dissociation and the sensitivity of the octamer to salt^{15,17}. Moreover, the role of the amino-terminal heptapeptide (region 1) in stabilizing the hydrophobic interaction cluster has been confirmed by several mutants in this region, all of which resulted in a shift of the equilibrium from the octameric towards the dimeric form²⁰.

Side and top views of the Mi_b -CK octamer are shown in Fig. 3c, d. The complex fits into a tetragonal cell of dimensions $93 \times 93 \times 86 \text{ \AA}$. A channel of diameter $20 \text{ \AA}$ extends along the four-fold axis through the octamer (Fig. 3d). The N termini of the monomers protrude into the channel at half height, whereas the C termini are located at the top and bottom four-fold faces. Dimensions and symmetry of the octamer are in excellent agreement with results from electron microscopy^{19,21,22}. It has also been shown that the octamers preferentially bind with their four-fold faces to the negatively charged phospholipid head groups of inner and outer mitochondrial membranes, mainly by electrostatic interaction^{21,23}. The C-terminal region Lys 360 to Lys 380 (M in Fig. 1a), located at the four-fold faces of the octamer, contains several positively charged amino acids. It is likely that some of the residues Lys 360, Lys 361, Lys 364, Lys 369, Arg 379 and Lys 380, and possibly also His 113 and Lys 110, are involved in the binding of Mi_b -CK octamers to naturally occurring cardiolipin of the mitochondrial inner membrane, and to the negatively charged head groups of lipids found in the outer membrane. The membrane-binding regions identified in the structure could explain the ability of Mi_b -CK to mediate the observed adhesion between inner and outer mitochondrial membranes²³, and make it plausible that Mi -CK, the only isoform in the CK system to form octamers, has evolved to serve this additional function as a structural protein. □

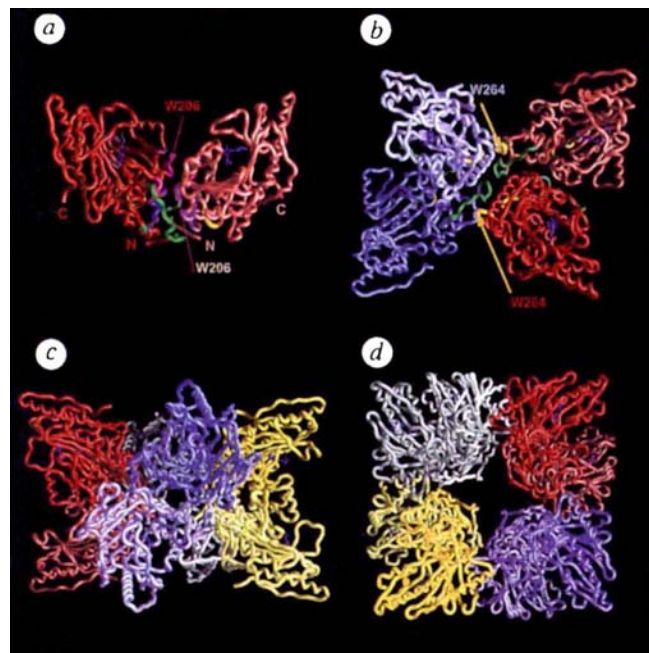


FIG. 3 Assembly of the Mi_b -CK octamer from monomers. The course of the polypeptide chains is shown in the 'worm' representation provided by the graphics program GRASP (A. Nicholls). The assembly involves four contact regions in each monomer, as defined in Fig. 1a. Regions 1–3 are used for dimer formation, and contacts 1 and 4 for the formation of octamers. The solvent-accessible areas³⁰ of a monomer, dimer and octamer are 17,057, 31,427 and $119,317 \text{ \AA}^2$, respectively. Thus dimer formation reduces the sum of the accessible areas of its monomers by $2,687 \text{ \AA}^2$, which is even larger than the contact area of $1,849 \text{ \AA}^2$ in the tight complex between actin and DNase I. The accessible area of the octamer, assembled from four dimers, is reduced by $6,391 \text{ \AA}^2$. The amino-acid residues involved in the interactions form a hydrophobic cluster which is shielded from solvent. a, Dimer formation. Two monomers, shown in red and light red, assemble into elongated dimers (length $92 \text{ \AA}$, width $42 \text{ \AA}$), which are shaped like bananas. The position of Trp 206 in contact region 3 is indicated. b, Assembly of two dimers. The position of Trp 264, a residue that is known to contribute to the stability of the octamer by contacting an N-terminal region 1 (green) of a neighbouring dimer, is indicated. c, Assembly of the Mi_b -CK octamer from four dimers. The viewing direction is roughly along a two-fold axis of the octamer. d, View of the Mi_b -CK octamer along its four-fold symmetry axis. A channel of diameter $20 \text{ \AA}$ extends through the whole octamer.

Received 13 November 1995; accepted 29 March 1996.

- Watts, D. C. *The Enzymes* 3rd edn (ed. Boyer, P. D.) 383–455 (Academic, New York, 1973).
- Kenyon, G. L. & Reed, G. H. *Adv. Enzymol.* **54**, 367–426 (1983).
- Wallmann, T., Wyss, M., Brdiczka, D., Nicolay, K. & Eppenberger, H. M. *Biochem. J.* **281**, 21–40 (1992).
- Hossle, J. P. et al. *Biochem. biophys. Res. Commun.* **151**, 408–416 (1988).
- Mühlebach, S. M. et al. *Molec. cell. Biochem.* **133/134**, 245–263 (1994).
- Taylor, S. S. & Radzio-Andzelm, E. *Structure* **2**, 345–355 (1994).
- Wyss, M., James, P., Schlegel, J. & Wallmann, T. *Biochemistry* **32**, 10727–10735 (1993).
- Olcott, M. C., Bradley, M. L. & Haley, B. E. *Biochemistry* **33**, 11935–11941 (1994).
- Buechter, D. D., Medzihradsky, K. F., Burlingame, A. L. & Kenyon, G. L. *J. Biol. Chem.* **267**, 2173–2178 (1992).
- Taylor, J. S., Leigh, J. S. & Cohn, M. *Proc. natn. Acad. Sci. U.S.A.* **64**, 219–226 (1969).
- Cook, P. F., Kenyon, G. L. & Cleland, W. W. *Biochemistry* **20**, 1204–1210 (1981).
- James, P., Wyss, M., Lutsenko, S., Wallmann, T. & Carafoli, E. *FEBS Lett.* **273**, 139–143 (1990).
- Rosevear, P. R., Desmeules, P., Kenyon, G. L. & Mildvan, A. S. *Biochemistry* **20**, 6155–6164 (1981).
- Vasák, M., Nagayama, K., Wüthrich, K., Mertens, M. L. & Kägi, J. H. R. *Biochemistry* **18**, 5050–5055 (1979).
- Gross, M., Furter-Graves, E. M., Wallmann, T., Eppenberger, H. M. & Furter, R. *Protein Sci.* **3**, 1058–1068 (1994).
- Schlegel, J. et al. *J. Biol. Chem.* **263**, 16942–16953 (1988).
- Gross, M. & Wallmann, T. *Biochemistry* **34**, 6660–6667 (1995).
- Milner-White, E. J. & Watts, D. C. *Biochem. J.* **122**, 727–740 (1971).
- Schnyder, T., Engel, A., Lustig, A. & Wallmann, T. *J. Biol. Chem.* **263**, 16954–16962 (1988).
- Kaldis, P., Furter, R. & Wallmann, T. *Biochemistry* **33**, 952–959 (1994).
- Schnyder, T., Cyrklaff, M., Fuchs, K. & Wallmann, T. *J. struct. Biol.* **112**, 136–147 (1994).
- Schnyder, T., Gross, H., Winkler, H., Eppenberger, H. M. & Wallmann, T. *J. Cell Biol.* **112**, 95–101 (1991).
- Rojas, M., Hovius, R., Demel, R. A., Nicolay, K. & Wallmann, T. *J. Biol. Chem.* **266**, 20290–20295 (1991).
- Wallmann, T., Schlösser, T. & Eppenberger, H. M. *J. Biol. Chem.* **259**, 5238–5246 (1984).
- Schnyder, T., Winkler, H., Gross, H., Eppenberger, H. M. & Wallmann, T. *J. Biol. Chem.* **266**, 5318–5322 (1991).
- Schnyder, T., Sargent, D. F., Richmond, T. J., Eppenberger, H. M. & Wallmann, T. *J. molec. Biol.* **216**, 809–812 (1990).
- Kabsch, W. *J. appl. Crystallogr.* **26**, 795–800 (1993).
- Brünger, A. T. *X-PLOR Version 3.1* (Yale University, New Haven, CT, 1992).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).

ACKNOWLEDGEMENTS. We thank K. C. Holmes and H. M. Eppenberger for support; I. Schlichting for discussions and help collecting synchrotron data; D. Madden for reading the manuscript; G. Eulefeld, W. Gebhard and I. Naundorf for technical assistance with figures and system programming; D. Sargent for help during the initial phase of structure analysis; and members of the CK group, especially M. Forstner, R. Furter, E. Furter, M. Gross, W. Hemmer and M. Wyss for contributions. This work was supported by the ETH-Zürich, Swiss National Science Foundation, H. Horten Foundation, Swiss Muscle Foundation, and Careal Holding AG.

CORRESPONDENCE and requests for materials should be addressed to W.K. (e-mail: kabsch@mpimf-heidelberg.mpg.de). The coordinates have been deposited in the Brookhaven Protein Database, accession number 1crk, where they will be held for one year before release.

ERRATUM

RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase

Andrew G. Polson, Brenda L. Bass & John L. Casey

Nature 380, 454–456 (1996)

DETAILS for correspondence regarding this work were incomplete as published: correspondence and requests for materials should be addressed to B.L.B. (on dsRAD) or to J.L.C. (on HDV). Their respective e-mail addresses are bass@tulip.med.utah.edu and caseyj@medlib.georgetown.edu. □

CORRECTION

Disruption of the nuclear hormone receptor ROR α in staggerer mice

Bruce A. Hamilton, Wayne N. Frankel, Anne W. Kerrebrock, Trevor L. Hawkins, William FitzHugh, Kenro Kusumi, Liane B. Russell, Ken L. Mueller, Victor van Berkel, Bruce W. Birren, Leonid Kruglyak & Eric S. Lander

Nature 379, 736–739 (1996)

THE labelling of Fig. 3a and b is incorrect: rostral is to the upper left and the fourth ventricle is seen as a triangular space above and to the right of the cerebellum. The ROR α 1 cDNA sequence has been submitted to GenBank, accession number U53228. □

KNOW YOUR COPY RIGHTS RESPECT OURS

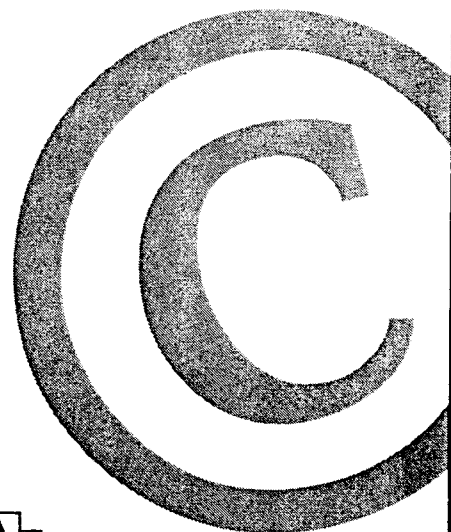
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Flexibility the key for neuroscientists

This feature looks at opportunities for neuroscientists in the United States and Europe. In the first review, Brendan Horton looks at some of the employment options for PhDs, many of whom are overspecialized for markets outside pure research. The rewards are high for those prepared to be flexible in their approach. The second article focuses on France, Germany and the United Kingdom. Quirin Schermeier examines prospects both in the traditional research areas and outside them, for those seeking relevant careers in an oversubscribed job market.

At the spring meeting of the Association of Neuroscience Programs and Departments, held at Georgetown University, scientific consultant Mathew Weinberg told the attendees "there is no better time to get a PhD". One might ask why, given the large number of faculty members keeping their positions well beyond 65 years of age and the lack of tenure-track positions becoming available. Funding is no longer as straightforward as it used to be, and even experienced researchers sometimes have difficulty securing funding.

In 1995 the National Academy of Sciences (NAS) published a report entitled *Reshaping the Graduate Education of Scientists and Engineers*, which states that "In the past, when students expected to become professors, graduate school was usually seen as a step on a simple career ladder. We are concerned that this concept is still held in some places. Departments should help students to regard their progress through graduate school as a journey with branches that require decisions."

Are there simply too many people with PhDs? The NAS report states that the percentage of people in academic employment 5 to 8 years after receiving a PhD in the

medical sciences in the United States decreased from 56.5 to 50.7 per cent between 1977 and 1991. In the biological sciences during the same time period, the

"It no longer holds that if you don't end up with a faculty research position that either the mentor or student has somehow failed"

number decreased from 59.1 to 45.7 per cent. By contrast, the percentage of scientists employed in business and industry increased in the medical sciences from 15.9 to 26.4 per cent, and in the biological sciences from 11.4 to 25.4 per cent during this period. Moreover, the report concludes that this trend is likely to continue, with most jobs for scientists and engineers in the coming years being created in business and industry. With more and more scientists with PhDs leaving the traditional career track, the question arises as to whether the traditional system is preparing students for the increasing variety of jobs that they may encounter during their careers. New oppor-

tunities are opening up all the time in previously unconsidered areas: are students taking advantage of them?

Mike Huerta, an associate director in the Division of Neuroscience and Behavioral Science at the National Institute of Mental Health (NIMH), says "we do not have too many PhDs, rather they are too narrowly trained". He feels this overspecialization has developed because graduate programmes have focused too much on producing tenure-track, faculty researchers. Neuroscientists need to be aware early in their education of the range of career options other than straight academic research. Knowing about the options may allow students to approach graduate education with more flexibility.

The NAS report notes that many employers in industry are concerned about the immediate suitability of graduates for entry-level jobs. The criticism is based in part on a belief that students have become too specialized for the demands of non-academic work.

Ivan Lieberburg, vice-president of research for Athena Neurosciences of South San Francisco, California, left the traditional academic career path after graduate school, medical training and a fully funded academic career as an associate professor because he felt it would allow him to make a more immediate impact. "Staying in academia would not have allowed me to tackle a problem as large as Alzheimer's or Parkinson's. We have large numbers of people working on this project, many more than could have been justified with the classic, academic funding mechanisms," says Lieberburg. These efforts have, for example, produced a diagnostic test for Alzheimer's disease.

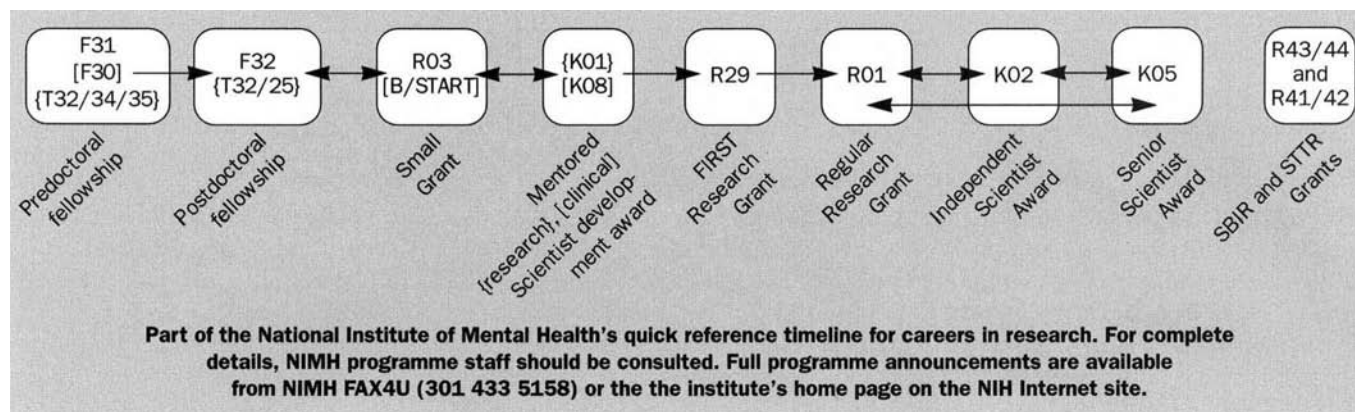
Skills and career guidance

Above all else, the report recommends the implementation of improved career guidance and information systems, so that students and their advisers can make informed career decisions based on current information. This, they feel, will lead to a more stable and more adaptive job market. The NAS does not advocate limits on PhD enrolment, but suggests that "the National Science Foundation should continue to improve the coverage, timeliness and

SOME OF THE MANY CAREER DEVELOPMENT RESOURCES ON THE INTERNET

Associations, organizations and companies

National Science Foundation	http://www.nsf.gov
National Academy of Science	www.nas.edu/
National Institutes of Health/	
National Institutes of Mental Health	www.nih.gov
Society for Neuroscience	www.sfn.org
Association of Neuroscience	
Departments and Programs	andp.physlog.uiowa.edu
Survival Skills and Ethics/	
Program University of Pittsburgh	www.pitt.edu/~survival/homepg.html
James. S. McDonnell Foundation	www.jsmf.org
Axon Instruments	www.axonet.com/career.html
BioOnline	www.bio.com
Wyeth-Ayerst Research	www.bio.com/hr/job/wyeth_1.html
Bristol-Meyers Squibb	www.bms.com
Eli Lilly	www.lilly.com
Genentech	www.gene.com
Nature	www.nature.com
Mental Health Net	www.cmhc.com
Research Biochemicals	www.callrbi.com



clarity of analysis of the data on the education and employment of scientists and engineers in order to support better national decision-making about human resources in science and technology".

Students can also improve their ability to adapt by participating in courses like the Survival Skills and Ethics Program, directed by Michael Zigmond of the Department of Neuroscience at the University of Pittsburgh. This course teaches graduate students and postdoctoral researchers skills that are often neglected in traditional graduate-school education. It operates locally at the university to teach members of the community professional survival skills, such as writing and speaking, learning and teaching, as well as 'career management'. The programme also integrates ethical issues into all aspects of the curriculum.

On the national level, 'trainer-of-trainers' workshops are designed to facilitate the development of programmes at other institutions. Workshop participants receive instruction and material to establish a course in survival skills and ethics at their own institution. The syllabus covers technical writing, grant management, making the graduate programmes widely accessible to minorities and the handicapped, oral presentations, attending meetings, relationships between student and faculty, and negotiating and managing a job. According to Zigmond, "there are now numerous viable career alternatives and we need to encourage individuals trained in science to broaden their definition of acceptable employment". These areas include scientific administration, industry, public policy, college teaching, journalism, publishing and science in law. "It no longer holds that if you don't end up with a faculty research position that either the mentor or the student has somehow failed".

Alternative funding sources

Various private organizations, such as the McDonnell Foundation in St Louis, Missouri, provide money for multidisciplinary studies. Susan Fitzpatrick, administrator of the McDonnell-Pew programmes in cognitive neuroscience, says that "its funding is intended to be 'investment capital' or 'seed money' for multidisciplinary programmes

that might fall through the cracks of the traditional funding process". Since 1990, the programme has provided US\$23 million in support of institutional centres and individual investigators. The programme's charter is to support innovative, interdisciplinary research that is unlikely to be funded from traditional sources.

The McDonnell Foundation is also monitoring the progress of those involved in its multidisciplinary and interdisciplinary programmes, and Fitzpatrick is due to report the survey's findings to the foundation within the year. In fields such as neuroimaging, researchers often find themselves between disciplines. The techniques and technologies of positron emission tomography (PET), magnetic resonance imaging (MRI) and magnetoencephalography (MEG) are useful to many, but people in this field do not have the same security or opportunity for faculty positions. Fitzpatrick believes that they often end up in radiology departments, providing a service for other departments rather than focusing on their own area of research.

Another area in which the neuroscientist can consider an alternative direction is in the development of small biotechnology businesses. Mike Huerta runs the Small Business Innovation Research (SBIR) and the Small Business Technology Transfer (STTR) programmes within NIMH. Each of the institutes and centres at the National Institutes of Health (NIH) have these programmes, which are designed to support research and development in small businesses with innovative technologies that have expected commercial potential. Both SBIR and STTR provide grant support to scientists at research institutions, colleges and universities. In financial year 1995, approximately US\$8 million was set aside for the SBIR programme within the NIMH. These grants are meant to establish contacts between researchers and those with small business interests. Information on these programmes is available in the NIMH area of the NIH web site (see table).

For those considering a career that integrates science and business, Weinberg recommends some study above and beyond the requirements of a PhD, such as a course in organizational behaviour and

communications. "People who work in the business world need a better understanding of organizational work and they cannot succeed without being able to communicate effectively".

There are also a variety of commonly overlooked areas where those with PhDs can look for opportunities off the traditional career track. For instance, many companies and investment banks that finance biomedical companies employ consultants to evaluate the quality of a scientific endeavour and help ventures proceed smoothly. According to the NAS report, Wall Street firms hire as many as 100 people with PhDs in science each year. Weinberg suggests that scientists may also find opportunities in the legal profession, not only as expert witnesses but as consultants to evaluate the scientific merit of evidence, help to devise strategies in litigation, and analyse and review scientific literature as part of an investigation. He also points out that none of these areas necessarily requires further formal qualifications.

Job hunting on the Internet

For those who are anxious to get behind a keyboard and 'pound the electronic pavement', there are a number of Internet resources that are useful (the searchable international job site in *Nature* is just one example). Users can search for available positions by country, subject, organization and position.

In the area of neuroscience, there is the Mental Health Net, which contains a listing of associations and organizations related to mental health, complete with addresses and telephone numbers. This site provides links to the home pages of many sites, a linked subject index that ranges from alcohol and substance abuse to statistics and research design. 'Job link' is an area within the site that has employment announcements and openings, as well as a bulletin board where a curriculum vitae can be deposited. This site should also be useful for tracking down organizations and finding details of their upcoming meetings.

The NAS Internet site offers a career planning centre, which is particularly useful for scientific researchers just embarking on their career. It includes a bulletin board for

questions or advice, an advice centre where users can sign up to have an online mentor, an employment and research-funding centre for searching available openings or tracking down financial support, and a resource centre for full-text, online reports, such as *Reshaping the Graduate Education of Scientist and Engineers*, and the new student guide from the NAS.

The McDonnell Foundation home page offers programme information, guidelines

for how to apply, and a listing of the grants and awards available from the foundation. The pertinent areas of funding are biomedical and behavioural sciences, including cognitive neuroscience, and research and innovation in education. Of particular interest will be the McDonnell-Pew programme in cognitive neurosciences, as well as a new programme that focuses on collaborative projects in cognitive rehabilitation research.

B.H.

The view from Europe

EVERY autumn, like migrating geese, thousands of European neuroscientists flock to the annual meeting of the US Society for Neuroscience. Why do they take the time and trouble to cross the Atlantic to submerge themselves in the bewildering crowds of tens of thousands of attendees?

The short answer is that, in contrast to Europe, neuroscience in the United States has a strong identity, a well-organized umbrella society, and its annual meeting acts as the world's largest market for making contacts in the hunt for collaborators or new employment.

Neuroscience in Europe has yet to settle on an identity of its own. The European Neuroscience Association (ENA) is hard pushed to attract 2,000 attendees, despite an estimated population of more than 12,000 neuroscientists. Things tend to happen more often at the national level, where travel to society meetings is cheaper, and scientists have the additional, often-overlooked, advantage of being able to relax and speak their own language. Leading neuroscientists are aware of the lack of cohesion at the European level. In an attempt to inspire political support, the European Commission was persuaded in 1990 to launch a Decade of the Brain initiative, to parallel the well-funded Decade of the Brain in the United States. The initiative was a damp squib, primarily because no new money was made available to support it.

An *ad hoc* task force of scientists was established to make recommendations for a special brain research programme in the fourth of the commission's 5-year framework programmes (1994–98), but in a burst of unrealistic optimism put forward plans that would have cost around 50 times the entire budget available for the commission's whole biomedicine programme, Biomed.

The proposal was duly rejected although, as a direct result of the initiative, a brain research programme covering basic and clinical research was included in the fourth Biomed framework programme. It receives 12 per cent of Biomed's ECU336 million (US\$415 million) fund, and has proved extremely popular: the first call for proposals early last year attracted 285

applications, mostly from the United Kingdom, France, Germany and Italy, of which 41 were accepted. Neuroscience now also has a special place in the commission's biotechnology programme and will receive 6 per cent of the funds, around ten times more than it received within the third Biotech framework programme.

Things could get even better. The European parliament last month called for a special programme for Alzheimer's disease to be established within the fourth framework programme, but again it is not clear whether or not extra money could be made available for it.

Despite its lack of cohesion, Europe has a very strong base in neurosciences and active national associations. In the United Kingdom, for example, the Brain Research Association last year published a booklet for secondary schools, aimed at encouraging students to apply to study neurosciences at university. The booklet, which schools can use almost as a textbook, has

been a big success, judging from the number of orders received, from as far afield as Australia and New Zealand. But whether it will be successful in increasing active interest in studying neuroscience is harder to assess. Neuroscience, a term coined only 50 years ago, is a diverse discipline, encompassing a wide range of subjects including cell and molecular biology, physiology, pharmacology, psychology, and even information science, and very few British universities offer an undergraduate course bearing the name 'neuroscience'.

This can be seen as both positive and negative, according to Gavin Swanson, editor of the Cambridge-based journal *Trends in Neurosciences*. On the one hand, because the subject is not entrenched in tradition, neuroscience remains flexible and "ideally placed to respond to market forces" he says. On the other hand, the lack of a strong identity could reduce its lobbying power for limited funds.

This latter point also worries Geoff Bennett, a lecturer at Nottingham University, who runs a group within the British Neuroscience Association to promote neuroscience teaching. But Rudolf Menzel, a professor of zoology at the Free University of Berlin, sees diversification as only an advantage, "because we can get money from a wide range of funding organizations and programmes".

To represent this diversity better, the Brain Research Association, which has 1,400 members, voted in March to change its name to the British Neuroscience Association (BNA) because many members felt the old name inappropriately excluded scientists working on other areas, for exam-

Initial progress in neuroimaging

In fields such as neuroimaging, researchers often find themselves providing a service to those in other disciplines. But these techniques are also advancing knowledge in their own right. The description of precise functional-anatomical correlations is producing new hypotheses about the functional architecture of the brain. In this image, a PET perfusion scan from an activation experiment designed to isolate mental components of upper limb movement is co-



registered with the subject's own anatomical MRI scan. Computer algorithms allow direct assessment of the precise cerebral structures involved. In this case, an imagined movement when compared to attending to the limb to be moved (preparing set), results in the opercular activations bilaterally (among other activations) even though the mental tasks are being directed specifically at the right upper limb. For the first time, such biological correlates of mental activity allow discussion of its nature in scientific terms rather than by introspections and speculation. (Image: Wellcome Dept of Cognitive Neurology, Institute of Neurology, London.)

ple spinal cord, peripheral nervous system and muscle.

Similarly, the German society Neurobiologentagung, which was founded in 1973 and attracts around 1,000 scientists to its informal annual meeting in Spring, has been felt to exclude clinically oriented scientists because its name confines it to neurobiologists. A new, more formal society was established in 1993, called the German Society for Neuroscience. It held its first meeting in Berlin in February this year. The head of the new society, Michael Frotscher, professor of anatomy at Freiburg University, says that its aim is to "support interaction in neuroscience, and have different subject sections, like the US Society for Neuroscience, which is our model". At present the German society has 1,000 members, predominantly molecular and cell biologists, but hopes to expand to include research-oriented clinicians in the near future. Most of the scientific proceedings take place in English.

By contrast, the language at the French society for neuroscience, which has around 2,000 members, from Belgium and French Canada as well as France, is unashamedly French. It has no cultural problem with its name: Société des Neurosciences.

Where is the work?

Most neuroscientists in Europe work in universities or publicly funded research institutes. Three-quarters of Germany's estimated 2,000 neuroscientists work in universities, the rest being located in Max Planck institutes (MPIs), such as the Institute for Biophysiological Chemistry in Göttingen, the Institute for Brain Research in Frankfurt, the Institute for Psychiatry near Munich, and the Institute for Biological Cybernetics in Tübingen. Others work at the so-called blue-list institutes, such as the new Institute for Neurobiology in Magdeburg, or at Fraunhofer applied research institutes, such as the Institute for Biomedical Technology in St Ingbert, which specializes in neural networks and neuronal electronics.

In France, an estimated population of 4,000 neuroscientists are evenly spread between universities and publicly funded research institutes (CNRS and INSERM), with a much smaller percentage working in industry. Public funding is therefore particularly important, and neuroscience research attracts reasonably generous funding, despite complaints from the academic community that it is not enough. In the United Kingdom, one of the largest funding bodies is the Wellcome Trust, which provides well over £320 million (US\$480 million) per year for 3-5-year project and programme grants in areas such as mental health, basic neuroscience and vision research. Neuroscientists have also benefited from the trust's new and generous 4-year Research Career Development Programme. Other significant funding agencies are the Medical Research Council and the Biotechnology and Biological Sci-

ences Research Council.

In France, CNRS and INSERM, the two most important research funding organizations, together provided FF150 million (US\$30 million) last year for neurosciences. "We are not at all happy with this", says Joël Bockaert, director of research at the CNRS-INSERM Institute for Pharmacology and Endocrinology in Montpellier, "because it is around 30 per cent less than what we received a few years ago". This mainly reflects the dramatic decrease in the general CNRS budget.

"Prospects are tough... there are more qualified neuroscientists than vacancies available"

In Germany, the Deutsche Forschungsgemeinschaft (DFG), the main organization for funding university research, last year gave DM95 million of its DM630 million (US\$ 420 million) biomedical sciences fund to neuroscience. The Max Planck Society's institutional budget for neuroscience's was DM45 million, and that of the blue-list institutions DM18 million.

It all sounds very generous, but not everyone is satisfied. Wolfgang Finger, head of biophysics at the University of Tübingen, and neuroscience advisor to the federal research ministry (the BMBF), says that the "combined DM160 million BMBF and DFG annual neuroscience funding must be placed in the context of the health costs, valued at over DM40 billion per year, for CNS diseases alone". Moreover, he says, although the general quality of basic research in neuroscience is very high in Germany, this cannot be said of applied research. "There is too little cooperation between universities and the pharmaceutical industry". This may be set to change. Future BMBF funding will shift emphasis towards applied research, he says. This will continue a trend begun a few years ago when the ministry commissioned a report on neurotechnology, which resulted in the launch last year of a five-year DM18 million interdisciplinary programme on retina implantation.

International funds are available to European neuroscientists not only from the European Union, but also from the Strasbourg-based Human Frontier Science Program (HFSP). The HFSP is supported by contributions from the governments of the European Union, Switzerland and the G7 countries, in particular Japan, and this year it will award over US\$7.5 million in the form of three-year research grants and 2-year fellowships in neuroscience. To promote the exchange of information and ideas, and to encourage further collaborative research, the HFSP recently initiated a series of workshops modelled loosely on the intensive study programmes that were

held in the 1960s as part of the neurosciences research programme organized by Francis Schmitt. The topic of the first workshop, held in 1995, was coincidence detection in neurons; another, later this year, will be on coding mechanisms in the dorsal and ventral streams of the cortex.

Opportunities in industry

What does industry have to offer neuroscientists in the way of employment? Very little in Germany, according to Henning Scheich, vice-president of the Blue-List Society and head of the Institute for Neurobiology. Major companies such as Hoechst and Bayer have shifted their main research departments to other countries where costs are lower and legislation, for example regarding the use of animals, is less restrictive. At Germany's other main pharmaceutical company, Schering, only ten neuroscientists are employed in its neuro-psychopharmacology team.

More can be expected from the Swiss giants Novartis, created this year from the merger of Ciba-Geigy and Sandoz, and Hoffmann-La Roche, which still have major research programmes in neural disorders despite shrinking research and development budgets. The Institute for Molecular Biology, which belongs to the British drug company Glaxo-Wellcome but is based in Geneva, has a core of 30 neuroscientists working on neurodegeneration.

The United Kingdom and France also have relatively good opportunities for industrial employment, particularly in drug development. In Britain, companies such as Sandoz, Merck Sharpe & Dohme, Astra, Glaxo-Wellcome, SmithKline Beecham and Parke-Davis have significant neuropharmacology programmes, many within university departments. In France, Synthelabo, Rhone-Poulenc and Sanofi are the three largest pharmaceutical companies, employing significant numbers of neuroscientists, but there are many smaller companies, including Servier, which recently opened large new research premises, and Pierre Fabre.

Neuroscience does not belong exclusively to neurobiologists, but also, more and more, to computer scientists. Industry is becoming much more interested in the development of neural networks to aid production processes, and large firms such as Siemens in Munich have teams of researchers dedicated to the development of 'intelligent machines'. There are also increasing numbers of smaller companies with interest in this area, such as Kratzer Automatisierung GmbH, also based in Munich.

Despite the range of potential opportunities for neuroscientists in Europe, employment prospects in the current recessionary times are not as bright as perhaps they should be. Says Bockaert: "prospects are tough and there are more qualified neuroscientists than positions available to use their talents".

Q.S.